Springer Series in Statistics

Dynamic Mixed Models for Familial Longitudinal Data

Springer

Springer Series in Statistics

Advisors:
P. Bickel, P. Diggle, S. Fienberg, U. Gather,
I. Olkin, S. Zeger

For other titles published in this series, go to
http://www.springer.com/series/692

Brajendra C. Sutradhar

Dynamic Mixed Models for Familial Longitudinal Data

Springer

Brajendra C. Sutradhar
Department of Mathematics and Statistics
Memorial University
A1C 5S7 Saint John's
Newfoundland and Labrador
Canada
bsutradh@mun.ca

ISSN 0172-7397
ISBN 978-1-4419-8341-1 e-ISBN 978-1-4419-8342-8
DOI 10.1007/978-1-4419-8342-8
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011921116

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Bhagawan Sri Sathya Sai Baba
my Guru, Mother and Father

**[Twameva Guru Cha Mata Twameva
Twameva Guru Cha Pita Twameva
Twameva Sarvam Mama Deva Deva]**

Preface

Discrete familial data consist of count or binary responses along with suitable covariates from the members of a large number of independent families, whereas discrete longitudinal data consist of similar responses and covariates collected repeatedly over a small period of time from a large number of independent individuals. As the statistical modelling of correlation structures especially for the discrete longitudinal data has not been easy, many researchers over the last two decades have used either certain 'working' models or mixed (familial) models for the analysis of discrete longitudinal data. Many books are also written reflecting these 'working' or mixed models based research. This book, however, presents a clear difference between the modelling of familial and longitudinal data. Parametric or semiparametric mixed models are used to analyze familial data, whereas parametric dynamic models are exploited to analyze the longitudinal data. Consequently, dynamic mixed models are used to analyze combined familial longitudinal data. Basic properties of the models are discussed in detail. As far as the inferences are concerned, various types of consistent estimators are considered, including simple ones based on method of moments, quasi-likelihood, and weighted least squares, and more efficient ones such as generalized quasi-likelihood estimators which account for the underlying familial and/or longitudinal correlation structure of the data. Special care is given to the mathematical derivation of the estimating equations.

The book is written for readers with a background knowledge of mathematics and statistics at the advanced undergraduate level. As a whole, the book contains eleven chapters including Chapters 2 and 3 on linear fixed and mixed models (for continuous data) with autocorrelated errors. The remaining chapters are also presented in a systematic fashion covering mixed models, longitudinal models, longitudinal mixed models, and familial longitudinal models, both for count and binary data. Furthermore, in almost every chapter, the inference methodologies have been illustrated by analyzing biomedical or econometric data from real life. Thus, the book is comprehensive in scope and treatment, suitable for a graduate course and further theoretical and/or applied research involving familial and longitudinal data.

Familial models for discrete count or binary data are generally known as the generalized linear mixed models (GLMMs). There is a long history on inferences in GLMMs with single or multiple random effects. In this GLMMs setup, the correlations among the responses under a family are clearly generated through the common random effects shared by the family members. However, as opposed to the GLMMs setup, it has not been easy to model the longitudinal correlations in generalized linear longitudinal models (GLLMs) setup. Chapter 1 provides an overview on difficulties and remedies with regard to (1) the consistent and efficient estimation in the GLMMs setup, and (2) the modelling of longitudinal correlations and subsequently efficient estimation of the parameters in GLLMs.

The primary purpose of this book is to present ideas for developing correlation models for discrete familial and/or longitudinal data, and obtaining consistent and efficient estimates for the parameters of such models. Nevertheless, in Chapter 2, we consider a clustered linear regression model with autocorrelated errors. There are two main reasons to deal with such linear models with autocorrelated errors. First, in practice, one may also need to analyze the continuous longitudinal data. Secondly, the knowledge of autocorrelation models for continuous repeated data should be helpful to distinguish them from similar autocorrelation models for discrete repeated data. Several estimation techniques, namely the method of moments (MM), ordinary least squares (OLS), and generalized least squares (GLS) methods are discussed. An overview on the relative efficiency performances of these approaches is also presented.

In Chapter 3, a linear mixed effects model with autocorrelated errors is considered for the analysis of clustered correlated continuous data, where the repeated responses in a cluster are also assumed to be influenced by a random cluster effect. A generalized quasi-likelihood (GQL) method, similar to but different from the GLS method, is used for the inferences in such a mixed effects model. The relative performance of this GQL approach to the so-called generalized method of moments (GMM), used mainly in the econometrics literature, is also discussed in the same chapter.

When the responses from the members of a given family are counts, and they are influenced by the same random family effect in addition to the covariates, they are routinely analyzed by fitting a familial model (i.e., GLMM) for count data. In this setup, the familial correlations among the responses of the members of the same family become the function of the regression parameters (effects of the covariates on the count responses) as well as the variance of the random effects. However, obtaining consistent and efficient estimates especially for the variance of the random effects has been proven to be difficult. With regard to this estimation issue, Chapter 4 discusses the advantages and the drawbacks of the existing highly competitive approaches, namely the method of moments, penalized quasi-likelihood (PQL), hierarchical likelihood (HL), and a generalized quasi-likelihood. The relatively new GQL approach appears to perform the best among these approaches, in obtaining consistent and efficient estimates for both regression parameters and the variance of the random effects (also known as the overdispersion parameter). This is demonstrated for the GLMMs for Poisson distribution based count data, first with single− and then with two-dimensional random effects in the linear predictor of the familial model. The aforementioned estimation approaches are discussed in detail in the parametric setup under the assumption that the random effects follow a Gaussian distribution. The estimation in the semiparametric and nonparametric set up is also discussed in brief.

Chapter 5 deals with familial models for binary data. These models are similar but different from those for count data discussed in Chapter 4. The difference lies in the fact that conditional on the random family effect, the distribution of the response of a member is assumed to follow the log-linear based Poisson distribution in the count data setup, whereas in the familial models for binary data, the response of a member is assumed to follow the so-called linear logistic model based binary distribution. This makes the computation of the unconditional likelihood and moments of the data more complicated under the binary set up as compared to the count data setup. A binomial approximation as well as a simulation approach is discussed to tackle this difficulty of integration over the distribution of the random effect to obtain unconditional likelihood or moments of the binary responses under a given family. Formulas for unconditional moments up to order four are clearly outlined for the purpose of obtaining the MM and GQL estimates for both regression and the overdispersion parameters.

In the longitudinal setup, the repeated responses collected from the same individual over a small period of time become correlated due to the influence of time itself. Thus, it is not reasonable to model these correlations through the common random effect of the individual. This becomes much clearer when it is understood that in some situations, conditional on the random effect, the repeated responses can be correlated. It has not, however, been easy to model the correlations of the repeated discrete such as count or binary responses. One of the main reasons for this is that unlike in the linear regression setup (Chapters 2 and 3), the correlations for the discrete data depend on the time-dependent covariates associated with the repeated responses. In fact, the modelling of the correlations for discrete data, even if the covariates are time independent, has also not been easy. Over the last two decades, many existing studies, consequently, have used arbitrary 'working' correlations structure to obtain efficient regression estimates as compared to the moment or least squares estimates. This is, however, known by now that this type of 'working' correlations model based estimates [usually referred to as the generalized estimating equations (GEE) based estimates] may be less efficient than the simpler moment or least squares estimates. Chapter 6 deals with a class of autocorrelation models constructed based on certain dynamic relationships among repeated count responses. When covariates are time independent, in this approach, it is not necessary to identify the true correlation structure for the purpose of estimation of the regression coefficients. A GQL approach is used which always produces consistent and highly efficient regression estimates, especially as compared to the moment or independence assumption based estimates. The modelling for correlations when covariates are time dependent is also discussed in detail. In order to use the GQL estimation approach, this chapter also demonstrates how to identify the true correlation structure of the data when it is assumed that the true model belongs to an autocorrelations class.

Similar to Chapter 6, Chapter 7 deals with dynamic models and various inference techniques including the GQL approach for the analysis of repeated binary data collected from a large number of independent individuals. Note that the correlated binary models based on linear dynamic conditional probabilities (LDCP) are quite different from those dynamic models discussed in Chapter 6 for the repeated count data. Furthermore, for the cases where it is appropriate to consider that the means and variances of repeated binary responses over time may maintain a recursive relationship, Chapter 7 provides a discussion on the inferences for such data by fitting a binary dynamic logit (BDL) model.

Chapter 8 develops a longitudinal mixed model for count data as a generalization of the longitudinal fixed effects model for count data discussed in Chapter 6. This generalization arises in practice because of the fact that if the response of an individual at a given time is influenced by the associated covariates as well as a random effect of the individual, then this random effect will remain the same throughout the data collection period over time. In such a situation, conditional on the random effect, the repeated responses will be influenced by the associated time dependent covariates as well as by time as a stochastic factor. Thus, conditional on the random effect, the repeated count responses will follow a dynamic model for count data as in Chapter 6. Note that unconditional correlations, consequently, will be affected by both the variance of the random effects as well as the correlation index parameter from the dynamic model. This extended correlation structure has been exploited to obtain the consistent and efficient GQL estimates for the regression parameters, as well as a consistent GQL estimate for the variance of the random effects.

By the same token as that of Chapter 8, Chapter 9 deals with various longitudinal mixed models for binary data. These models are developed based on the assumption that conditional on the individual's random effect, the repeated binary responses either follow the LDCP or BDL models as in Chapter 7. Conditional on the random effects, a binary dynamic probit (BDP) model is also considered. This generalized model is referred to as the binary dynamic mixed probit (BDMP) model. In general, the GQL estimation approach is used for the inferences. The GMM and maximum likelihood (ML) estimation approaches are also discussed.

Chapter 10 is devoted to the inferences in familial longitudinal models for count data. These models are developed by combining the familial models for count data discussed in Chapter 4 and the longitudinal models (GLLMs) for count data discussed in Chapter 6. The combined model has been referred to as the GLLMM (generalized linear longitudinal mixed model). In this setup, the count responses are two-way correlated, familial correlations occur due to the same random family effect shared by the members of a given family, and the longitudinal correlations arise due to the possible dynamic relationship among the repeated responses of a given member of the family. These two-way correlations are taken into account to develop the GQL estimating equations for the regression effects and variance component for the random family effects, and the moment estimating equation for the longitudinal correlation index parameter.

Chapter 11 discusses the inferences in GLLMMs for binary data. A variety of longitudinal correlation models is considered, whereas the familial correlations are developed through the introduction of the random family effects only. The GQL approach is discussed in detail for the estimation of the parameters of the models. Because the likelihood estimation is manageable when longitudinal correlations are introduced through dynamic logit models, this chapter, similar to Chapter 9, discusses the ML estimation as well. As a further generalization, two-dimensional random family effects are also considered in the dynamic logit relationship based familial longitudinal models. Both GQL and ML approaches are given for the estimation of the parameters of such multidimensional random effects based familial longitudinal models.

Acknowledgements

Apart from my own research with familial and longitudinal data over the last fifteen years, this book has benefitted tremendously from my joint research with Drs. Kalyan Das, R. Prabhakar Rao, Vandna Jowaheer, Gary Sneddon, Pranesh Kumar, and Patrick Farrell, among others. I fondly remember their individual research journey with me and thankfully acknowledge their contributions that helped me to reach a stage that I thought appropriate to initiate the writing of this book.

The presentation of the materials to cover the wide field of inferences for familial and longitudinal discrete data has not been easy. This presentation task has benefitted from my experience in teaching and guiding graduate students in this area over the last ten years. I am thankful to those students who made me think about the necessity of a book at their level in this familial and longitudinal setup.

I wish to sincerely thank my colleague Dr. Gary Sneddon for his comments on some of the chapters of the book, and my collaborator Dr. Vandna Jowaheer from the University of Mauritius, for reading almost the entire book with love and patience, and providing valuable remarks and suggestions. I also thankfully acknowledge the inspirational comments and suggestions from five anonymous referees at different stages during the preparation of the book.

It has been a pleasure to work with John Kimmel, Marc Strauss, and Matthew Amboy of Springer-Verlag. I also wish to thank the copy-editor, Ms. Valerie T. Greco, for her attention to detail and superb accuracy on the copy-edit.

The writing of this book would not have been possible without the support of my family. I would like to thank my wife Malina Sutradhar for her inspiration and endless care, and my daughter Rinku and son Subir for their constant encouragement during the preparation of this book. I would also like to give many thanks to my adorable granddaughter Riya and her parents Rinku and Meghal for the countless number of pleasant breaks over Skype video calls, which allowed me to re-energize and continue writing the book enthusiastically during the last year.

Contents

Chapter 1
Introduction

Discrete data analysis such as count or binary clustered data analysis has been an important research topic over the last three decades. In general, two types of clusters are frequently encountered. First, a cluster may be formed with the responses along with associated covariates from the members of a group/family. These clustered responses are supposed to be correlated as the members of the cluster share a common random group/family effect. In this book, we refer to this type of correlation among the responses of members of same family as the familial correlation. Second, a cluster may be formed with the repeated responses along with associated covariates collected from an individual. These repeated responses from the same individual are also supposed to be correlated as there may be a dynamic relationship between the present and past responses. In this book, we refer to these correlations among the repeated responses collected from the same individual as the longitudinal correlations. It is of interest to fit a suitable parametric or semi-parametric familial and/or longitudinal correlation model primarily to analyze the means and variances of the data. Note that the familial and longitudinal correlations, however, play an important role in a respective setup to analyze the means and variances of the data efficiently.

1.1 Background of Familial Models

There is a long history of count and binary data analysis in the familial setup. It is standard to consider that a count response may be generated by a Poisson distribution based log linear model [Nelder (1974), Haberman (1974), and Plackett (1981)]. Similarly, a binary response may be generated following a linear logistic model [Berkson (1944, 1951), Dyke and Patterson (1952), and Armitage (1971)]. Because both Poisson and binary distributions belong to a one-parameter exponential family, both log linear and linear logistic models belong to the exponential family based generalized linear models (GLMs) [McCullagh and Nelder (Section 2, 1983)]. Consequently, when the count or binary responses from the members of a

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_1,

family form a cluster, a generalized linear mixed model (GLMM) is used to analyze such family based cluster data, where GLMMs are generated from the GLMs by adding random effects to the so-called linear predictor. Under the assumption that these random effects follow the Gaussian distribution, many authors such as Schall (1991), Breslow and Clayton (1993), Waclawiw and Liang (1993), Breslow and Lin (1995), Kuk (1995), Lee and Nelder (1996) [see also Lee and Nelder (2001)], Sutradhar and Qu (1998), Jiang (1998), Jiang and Zhang (2001), Sutradhar and Rao (2003), Sutradhar (2004), Sutradhar and Mukerjee (2005), Jowaheer, Sutradhar, and Sneddon (2009), and Chowdhury and Sutradhar (2009) have studied the inferences in GLMMs mainly for the consistent estimation of both regression effects of the covariates on the responses and the variance of the random effects. Note that in the familial, i.e., in GLMM set up, the variance of the random effects is in fact the familial correlation index parameter, which is not so easy to estimate consistently.

Schall (1991) and Breslow and Clayton (1993), among others, have used a best linear unbiased prediction (BLUP) analogue estimation approach, where random family effects are treated to be the fixed effects [Henderson (1963)] and the regression and variance components of the GLMMs are estimated based on the so-called estimates of the random effects. Waclawiw and Liang (1993) have developed an estimating function based approach to component estimation in the GLMMs. In their approach they utilize the so-called Stein-type estimating functions (SEF) to estimate both the random effects and their variance components. In connection with a Poisson mixed model with a single component of dispersion, Sutradhar and Qu (1998) have, however, shown that the so-called SEF approach of Waclawiw and Liang (1993) never produces consistent estimates for the variance component of the random effects, whereas the BLUP analogue approach of Breslow and Clayton (1993) may or may not yield a consistent estimate for the variance of the random effects (also known as the overdispersion parameter), which depends on the cluster size and the associated design matrix. In order to remove biases in the estimates, Kuk (1995) and Lin and Breslow (1996), among others, provided certain asymptotic bias corrections both for the regression and the variance component estimates. But, as Breslow and Lin (1995, p. 90) have shown in the context of binary GLMM with a single component of dispersion that the bias corrections appear to improve the asymptotic performance of the uncorrected quantities only when the true variance component is small, more specifically, less than or equal to 0.25.

As opposed to the BLUP analogue approach of Breslow and Clayton (1993) (also known as the so-called penalized quasi-likelihood (PQL) approach), Jiang (1998) proposed a simulated moment approach that always yields consistent estimators for the parameters of the mixed model. The moment estimators may, however, be inefficient. In the context of the binary mixed model, Sutradhar and Mukerjee (2005) have introduced a simulated likelihood approach which produces more efficient estimates than the simulated moment approach of Jiang (1998). To overcome the inefficiency of the moment approach, Jiang and Zhang (2001) have suggested an improvement over the method of moments. It, however, follows from Sutradar (2004) that the estimators obtained based on the improved method of moments (IMM) may also be highly inefficient as compared to the estimators obtained based on a generalized quasi-likelihood (GQL) approach. The GQL estimators are consistent and highly efficient, the exact maximum likelihood estimators being fully efficient (i.e., optimal) which are, however, known to be cumbersome to compute. In particular, the estimation of the variances of the estimators by the maximum likelihood approach may be extremely difficult (Sutradhar and Qu (1998)).

Lee and Nelder (1996) have suggested hierarchical likelihood (HL) inferences for the parameters in GLMMs. This HL approach is similar to but different from the PQL approach of Breslow and Clayton (1993). They are similar as in both approaches the estimation of the regression effects and the variance of the random effects is done through the prediction of the random effects by pretending that the random effects are fixed parameters even though they are truly unobservable random effects. To be specific, in the first step, both PQL and HL approaches estimate the regression parameters and the random effects. The difference between the two approaches is that the PQL approach estimates them by maximizing a penalized quasi-likelihood function, whereas the HL approach maximizes a hierarchical likelihood function. In the second step, in estimating the variance of the random effects, the PQL approach maximizes a profile quasi-likelihood function, whereas the HL approach maximizes an adjusted profile hierarchical likelihood function. Consequently, the HL approach may also suffer from similar inconsistency problems due to similar reasons that cause inconsistency in the PQL approach. This is also evident from Chowdhury and Sutradhar (2009) where it is shown in the context of a Poisson mixed model with a single random effect that the HL approach appears to produce highly biased estimates for the regression parameters, especially when the variance of the random family effects is large. The biases of the HL estimates also appear to vary depending on the cluster/family sizes. These authors have further demonstrated that the GQL approach [Sutradhar (2004)] produces almost unbiased and consistent estimates for all parameters of the Poisson mixed model irrespective of the cluster size and the magnitude of the variance of the random effects. In the context of Poisson mixed models with two variance components, Jowaheer, Sutradhar, and Sneddon (2009) have shown that the GQL approach performs very well in estimating the parameters of this larger mixed model. In this book, among other estimation approaches, we exploit this GQL approach for the estimation of the parameters both in count and binary mixed models. The GQL approach produces consistent as well as highly efficient estimates as compared to other competitive approaches such as moment, PQL, and HL estimation approaches.

1.2 Background of Longitudinal Models

In the longitudinal setup, a small number of repeated responses along with a set of covariates are collected from a large number of independent individuals over the same time points within a small period of time. Note that irrespective of the situations whether one deals with count or binary data, it is most likely that the repeated responses will be autocorrelated. Furthermore, these autocorrelations will exhibit stationary pattern [Sutradhar (2003,2010)] when the covariates collected over time from an individual are time independent. If the covariates are, however, time dependent, then the correlations will exhibit a nonstationary pattern [Sutradhar (2010)]. But it is not easy to write either a probability model or a correlation model for the repeated count and binary responses, even if the covariates are time independent (stationary correlations case). For the nonstationary cases, the construction of the probability or correlation models will be much more complicated.

Many authors including Liang and Zeger (1986) have used a 'working' stationary correlation structure based generalized estimating equation (GEE) approach for the estimation of the regression effects, even though the repeated data are supposed to follow a nonstationary correlation structure due to time-dependent covariates. This GEE approach, directly or indirectly, has also been incorporated in many research monographs or textbooks. For example, one may refer to Diggle et al (2002), and Molenberghs and Verbeke (2005). However, as demonstrated by Crowder (1995), because of the uncertainty of definition of the working correlation matrix, the Liang−Zeger approach may in some cases lead to a complete breakdown of the estimation of the regression parameters. Furthermore, Sutradhar and Das (1999) have demonstrated that even though the GEE approach in many situations yields consistent estimators for the regression parameters, this GEE approach may, however, produce less efficient estimates than the independence assumption based quasi-likelihood (QL) or moment estimates. These latter QL or moment estimates are also 'working' independence assumption based GEE estimates. Note that for the purpose of a demonstration on efficiency loss by the GEE approach, Sutradhar and Das (1999), similar to Liang and Zeger (1986), have considered the stationary correlation structure in the context of longitudinal binary data analysis even though the covariates were time dependent. In fact the use of a 'working' stationary correlation matrix in place of the true stationary correlation matrix may also produce less efficient estimates than the 'working' independence assumption based GEE or QL or moment estimates. This latter situation is demonstrated by Sutradhar (2010, Section 3.1) through an asymptotic efficiency comparison for stationary repeated count data. These studies by Crowder (1995), Sutradhar and Das (1999), Sutradhar (2003), and Sutradhar (2010) reveal that the GEE approach cannot be trusted for the regression estimation for the discrete such as longitudinal binary or count data.

Fitzmaurice, Laird and Rotnitzky [1993, eqns (2)–(4)] discuss a GEE approach following Liang and Zeger (1986) but estimate the 'working' correlations through a second set of estimating equations which is quite similar to the set of estimating equations for the regression parameters. Note that in this approach, the construction of the estimating equations for the 'working' correlation parameters requires another 'working' correlation matrix consisting of the third− and fourth-order moments of the responses, although Fitzmaurice et al (1993) use a 'working' independence approach to construct such higher-order moments based estimating equations. Similar to Fitzmaurice et al (1993), Hall and Severini (1998) also estimate the regression and the 'working' correlation parameters simultaneously. Hall and Severini (1998) referred to their approach as the extended generalized estimating equations

(EGEE) approach. This EGEE approach, unlike the approach of Fitzmaurice, Laird and Rotnitzky (1993) does not require any third− and fourth-order moments based estimating equations for the 'working' correlation parameters. It rather uses a set of second-order moments based estimating equations for the 'working' correlation parameters. Note however that these GEE based approaches of Fitzmaurice, Laird and Rotnitzky (1993) and Hall and Severini (1998) also cannot be trusted for the same reasons that the GEE cannot be trusted. We refer to Sutradhar (2003) and Sutradhar and Kumar (2001) for details on the inefficiency problems encountered by the aforementioned extended GEE approaches.

As a resolution to this inference problem for consistent and efficient estimation of the regression effects in the longitudinal setup, Sutradhar (2003, Section 3) has suggested an efficient GQL approach, which does not require the identification of the underlying autocorrelation structure, provided the covariates are time independent. This GQL approach for the discrete correlated data is in fact an extension of the QL approach (or weighted least squares approach) for the independent data introduced by Wedderburn (1974), among others. Sutradhar (2010) has introduced nonstationary autocorrelation structures for the cases when covariates are time dependent, and applied the GQL approach for consistent and efficient estimation of the regression effects. Sutradhar (2010) has also provided an identification of the autocorrelation technique for the purpose of the construction of an appropriate GQL estimating equation. In this book, we have exploited this GQL approach for the estimation of the parameters both in a longitudinal and familial setup.

Zhao and Prentice (1990), Prentice and Zhao (1991), and Zhao, Prentice, and Self (1992) have described extensions of the GEE methodology to allow for joint estimation of the regression and the true longitudinal correlation parameters in a binary longitudinal model. More specifically, Zhao and Prentice (1990) propose a joint probability model that is based on the 'quadratic exponential family,' with the three− and higher-way association parameters equal to zero. The 'quadratic exponential family' based association parameters are then estimated by using the likelihood estimating or equivalently, the generalized estimating equations approach. Similarly, a partly exponential model is introduced by Zhao, Prentice, and Self (1992) which accommodates the association between the responses, and the likelihood or equivalently the GEE approach was used to estimate the mean and the association parameters of the model. These GEE based methods for the joint estimation are referred to as the GEE2 approaches. Some of these GEE2 approaches, however, encounter convergence problems especially for the estimation of the longitudinal correlations [Sutradhar (2003)].

For continuous longitudinal data, some authors, for example, Pearson et al. (1994), Verbeke and Molenberghs (2000, Chapter 3), and Verbeke and Lesaffre (1999), modelled the means of the repeated responses as a linear or quadratic function over time. In this approach, time is considered to be a deterministic factor and hence times do not play any role to correlate the responses. Diggle, Liang, and Zeger (1994) [see also Diggle et al (2002), Verbeke and Molenberghs (2000, Chapter 3)] argue that the effect of serial (lag) correlations is very often dominated by suitable random effects and hence they modelled the longitudinal correlations through the introduction of the random effects. However, contrary to the above argument, it follows, for example, from Sneddon and Sutradhar (2004) that even though the random effects generate an equicorrelation structure for the repeated responses, they do not appear to address the time effects. This is because these individual specific random effects may remain the same throughout the data collection period and hence cannot represent any time effects. For this reason, Sneddon and Sutradhar (2004) modelled the longitudinal correlations of the responses through the autocorrelation structure of the errors involved in a linear model.

Similar to the continuous longitudinal setup, some authors have modelled the correlations of the repeated discrete data through the introduction of the time-specific random effects in the conditional mean functions of the data. For example, similar to GLMMs, Thall and Vail (1990) [see also Heagerty (1999) and Neuhaus (1993)] modelled the correlations of the repeated count data with overdispersion through the introduction of the random effects. However, one of the problems with this type of approach is that the lag correlations of the repeated responses in a cluster may become complicated. Furthermore, as argued by Jowaheer and Sutradhar (2002), this approach is unable to generate any pattern such as Gaussian type autocorrelation structure among responses as alluded in Liang and Zeger (1986), for example. In this book, following Sutradhar (2003, 2010), we have emphasized a class of Gaussian type autocorrelation structures to model the longitudinal correlations for both count and binary data. The random effects are used to model the overdispersion and/or familial correlations.

References

1. Armitage, P. (1971). *Statistical Methods in Medical Research.* Oxford: Blackwell.
2. Berkson, J. (1944). Application of the logistic function to bio-assay. *J. Am. Statist. Assoc.,* 39, 357 – 365.
3. Berkson, J. (1951). Why I prefer logits to probits. *Biometrics,* 7, 327 – 339.
4. Breslow, N, E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. . *J. Amer. Stat. Assoc.,* 88, 9 – 25.
5. Breslow, N. E. & Lin, X. (1995). Bias correction in generalized linear models with a single component of dispersion. *Biometrika,* 82, 81 – 92.
6. Chowdhury, M. R. I. & Sutradhar, B. C. (2009). Generalized quasilikelihood versus hierarchial likelihood inferences in generalized linear mixed models for count data. *Sankhya B: Indian J. Stat.,* 71, 55 – 78.
7. Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika,* 82, 407 – 410.
8. Diggle, P. J., Heagety, P., Liang, K.-Y., & Zeger, S. L. (2002). *Analysis of Longitudinal Data.* Oxford Science. Oxford: Clarendon Press.
9. Diggle, P. J., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford Science. Oxford: Clarendon Press.
10. Dyke, G. V. & Patterson, H. D. (1952). Analysis of factorial arrangements when data are proportion. *Biometrics,* 8, 1 – 12.
11. Fitzmaurice, G. M., Laird, N. M., & Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses. *Statist. Sci.,* 8, 284 – 309.
12. Haberman, S. J. (1974). Log-linear models for frequency tables with ordered classifications. *Biometrics,* 36, 589 – 600.

13. Hall, D. B., & Severini, T. A. (1998). Extended generalized estimating equations for clustered data. *J. Amer. Statist. Assoc.,* 93, 1365 – 75.
14. Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics,* 55, 688 – 98.
15. Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding,* National Research Council Publication No. 892, pp. 141 – 63. National Academy of Sciences.
16. Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.,* 93, 720 – 729.
17. Jiang, J. & Zhang, W. (2001). Robust estimation in generalized linear mixed models. *Biometrika,* 88, 753 – –765.
18. Jowaheer, V., Sutradhar, B. C. & Sneddon, G. (2009). On familial Poisson mixed models with multi-dimensional random effects. *J. Statist. Comp. Simul.,* 79, 1043 – 1062.
19. Jowaheer, V. & Sutradhar, B. C. (2002). Analysing longitudinal count data with overdispersion. *Biometrika,* 89, 389 – 399.
20. Kuk, A. Y. C. (1995). Asymptotically unbised estimation in generalized linear models with random effects. *J. R. Stastist. Soc. B,* 58, 619 – 678.
21. Lee, Y. & Nelder, J.A. (1996). Hierarchical generalized linear models. *J. R. Statist. Soc. B,*, 58, 619 – –678.
22. Lee, Y. & Nelder, J.A. (2001). Hierarchical generalized linear models: A synthesis of generalized linear models, random-effect models and structured dispersions. *Biometrika,*, 88, 987 – –1006.
23. Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika,* 73, 13 – 22.
24. Lin, X. & N. E. Breslow (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Am. Statist. Assoc.,* 91, 1007 – 1016.
25. McCullagh, P. & J. A. Nelder (1983, 1989). *Generalized Linear Models.* Chapman and Hall, London.
26. Molenberghs, G. & Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* Springer, New York.
27. Nelder, J. A. (1974). Log-linear models for contingency tables: A generalization of classical least squares. *Appl. Statist.,* 23, 323 – 329.
28. Neuhaus, J. M. (1993). Estimation efficiency and tests of covariate effects with clustered binary data. *Biometrics,* 49, 989 – 996.
29. Pearson, J. D., Morrell, C. H., Landis, P. K., Carter, H. B., & Brant, L. J. (1994). Mixed-effects regression models for studying the natural history of prostate disease. *Statist. Med.,* 13, 587 – 601.
30. Plackett, R. L. (1981). *The Analysis of Categorical Data.* Griffin, London.
31. Prentice, R. L. & Zhao, L.P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics,* 47, 825 – 39.
32. Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika,* 78, 719 – 27.
33. Sneddon, G. & Sutradhar, B. C. (2004). On semi-parametric familial-longitudinal models. *Statist. Probab. Lett.,* 69, 369 – 379.
34. Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statist. Sci.,* 18, 377 – 93.
35. Sutradhar, B.C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhyā: Indian J. Statist.,* 66, 261 – –289.
36. Sutradhar, B. C. (2010). Inferences in generalized linear longitudinal mixed models. *Canad. J. of Statist.,* 38, 174 – 196.
37. Sutradhar, B. C. & Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika,* 86, 459 – 65.
38. Sutradhar, B. C. & Kumar, P. (2001). On the efficiency of extended generalized estimating equation approaches. *Statist. Probab. Lett.,* 55, 53 – 61.

39. Sutradhar, B. C. & Mukerjee, R. (2005). On likelihood inference in binary mixed model with an application to COPD data. *Comput. Statist. Data Anal.,* 48, 345 – 361.
40. Sutradhar, B. C. & Z. Qu (1998). On approximate likelihood inference in Poisson mixed model. *Canad. J. Statist.,* 26, 169 – 186.
41. Sutradhar, B. C. & Rao, R. P. (2003). On quasi-likelihood inference in generalized linear mixed models with two components of dispersion. *Canad. J. Statist.,* 31, 415 – 435.
42. Thall, P. F. & Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics,* 46, 657 – 71.
43. Verbeke, G. & Lesaffre, E. (1999). The effect of drop-out on the efficiency of longitudinal experiments. *Appl. Statist.,* 48, 363 – 375.
44. Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* Springer, New York.
45. Waclawiw, M. A. & K.-Y. Liang (1993). Prediction of random effects in the generalized linear model. *J. Am. Statist. Assoc.,* 88, 171 – 178.
46. Wedderburn, R. (1974). Quasilikelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika,* 61, 439 – –447.
47. Zhao, L. P. & Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika,* 77, 642 – 48.
48. Zhao, L. P., Prentice, R. L. & Self, S. G. (1992). Multivariate mean parameter estimation by using a partly exponential model. *J. Roy. Statist. Soc. Ser. B,* 54, 805 – 811.

Chapter 2
Overview of Linear Fixed Models for Longitudinal Data

In a longitudinal setup, a small number of repeated responses along with certain multidimensional covariates are collected from a large number of independent individuals. Let $y_{i1}, \ldots, y_{it}, \ldots, y_{iT_i}$ be $T_i \geq 2$ repeated responses collected from the ith individual, for $i = 1, \ldots, K$, where $K \rightarrow \infty$. Furthermore, let $x_{it} = (x_{it1}, \ldots, x_{itp})'$ be the p-dimensional covariate vector corresponding to y_{it}, and β denote the effects of the components of x_{it} on y_{it}. For example, in a biomedical study, to examine the effects of two treatments and other possible covariates on blood pressure, the physician may collect blood pressure for $T_i = T = 10$ times from $K = 200$ independent subjects. Here the treatment covariate may be denoted by $x_{it1} = 1$, if the ith individual is treated by say treatment A, and $x_{it1} = 0$, if the individual is treated by the second treatment B. Let x_{it2}, x_{it3}, x_{it4}, and x_{it5}, respectively, denote the gender, age, smoking, and drinking habits of the ith individual. Thus, $p = 5$, and β denote the five-dimensional vector of regression parameters. Note that because $y_{i1}, \ldots, y_{it}, \ldots, y_{iT_i}$ are T_i repeated blood pressure collected from the same ith individual, it is likely that they will be correlated. Let $\Sigma_i = (\sigma_{iut})$ denote the $T_i \times T_i$ possibly unknown covariance matrix of these repeated responses. This type of correlated data is usually modelled by using the linear relationship

$$y_i = X_i\beta + \varepsilon_i, \tag{2.1}$$

where

$$y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT_i})'$$

is the vector of repeated responses,

$$X_i' = [x_{i1}, \ldots, x_{iT_i}]$$

is the $p \times T_i$ matrix of covariates for the ith individual, and

$$\varepsilon_i = [\varepsilon_{i1}, \ldots, \varepsilon_{it}, \ldots, \varepsilon_{iT_i}]'$$

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_2,

is the T_i-dimensional residual vector such that for all $i = 1, \ldots, K$, ε_i are independently distributed (id) with 0 mean vector and covariance matrix Σ_i. That is,

$$\varepsilon_i \overset{\text{id}}{\sim} (0, \Sigma_i).$$

It is of scientific interest to estimate β consistently and as efficiently as possible.

Note that even if the covariates are time dependent, in the present linear model setup, the residual vector ε_i is likely to have a stationary covariance structure. But, it is most likely that this structure belongs to a suitable class of stationary autocorrelation models such as autoregressive moving average models of order $q = 0, 1, 2, \ldots$ and $r = 0, 1, 2, \ldots$ [ARMA(q,r)] [Box and Jenkins (1970, Chapter 3)] or perhaps completely unknown. Further note that even though the residual covariance matrices for all $i = 1, \ldots, K$ are likely to have a common structure, their dimension will, however, be different for the unbalanced data. For this reason, one may denote the common covariance matrix by Σ, that is, $\Sigma_i = \Sigma$, only when $T_i = T$, for all $i = 1, \ldots, K$. In the longitudinal setup, it is convenient in general to express the covariance matrix Σ_i as

$$\begin{aligned} \Sigma_i &= (\sigma_{iut}) \\ &= A_i^{1/2} C_i A_i^{1/2}, \end{aligned} \tag{2.2}$$

where $A_i = \text{diag}[\sigma_{i11}, \ldots, \sigma_{itt}, \ldots, \sigma_{iT_iT_i}]$ and C_i is the $T_i \times T_i$ correlation matrix of $y_i = [y_{i1}, \ldots, y_{it}, \ldots, y_{iT_i}]'$. Note that, if $\sigma_{itt} = \text{var}(Y_{it}) = \sigma^2$ for all $t = 1, \ldots, T_i$ and the repeated responses are assumed to be independent (which is unlikely to hold in practice) i.e., $C_i = I_{T_i}$, a $T_i \times T_i$ identity matrix, then Σ_i reduces to

$$\Sigma_i = \sigma^2 I_{T_i}. \tag{2.3}$$

2.1 Estimation of β

2.1.1 Method of Moments (MM)

Irrespective of the cases whether the repeated responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT_i}$ are independent or correlated, one may always obtain the moment estimate of β by solving the moment equation

$$\sum_{i=1}^{K} [X_i'(y_i - X_i\beta)] = 0. \tag{2.4}$$

Let the moment estimator of β, the root of the moment equation (2.4), be denoted by $\hat{\beta}_M$. It is clear that $\hat{\beta}_M$ is easily obtained as

$$\hat{\beta}_M = \left[\sum_{i=1}^{K} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' y_i\right]. \tag{2.5}$$

Because $E[Y_i] = X_i\beta$ by (2.1), for a small or large K, it follows that $\hat{\beta}_M$ is unbiased for β, that is, $E[\hat{\beta}_M] = \beta$, with its covariance matrix given by

$$\begin{aligned} \text{cov}[\hat{\beta}_M] &= V_M \\ &= \left[\sum_{i=1}^{K} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' \Sigma_i X_i\right] \left[\sum_{i=1}^{K} X_i' X_i\right]^{-1}, \end{aligned} \tag{2.6}$$

where Σ_i is the covariance matrix of y_i, which may be unknown.

Note that when K is sufficiently large, it follows from (2.5) by using the multivariate central limit theorem [see Mardia, Kent and Bibby (1979, p. 51), for example] that $\hat{\beta}_M$ has asymptotically $(K \to \infty)$ a multivariate Gaussian distribution with zero mean vector and covariance matrix V_M as in (2.6). Note that in this large sample case, the covariance matrix V_M may be estimated consistently by using the sandwich type estimator

$$\hat{V}_M = \text{limit}_{K\to\infty} \left[\sum_{i=1}^{K} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' (y_i - \mu_i)(y_i - \mu_i)' X_i\right] \left[\sum_{i=1}^{K} X_i' X_i\right]^{-1}, \tag{2.7}$$

where $\mu_i = X_i\beta$ is known by using $\beta = \hat{\beta}_M$ from (2.5).

2.1.2 Ordinary Least Squares (OLS) Method

In this approach, the correlations among the repeated responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT_i}$, are ignored, and the ordinary least squares (OLS) estimator, say $\hat{\beta}_{OLS}$, of the regression parameter β in (2.1) is obtained by minimizing the sum of squared residuals

$$\begin{aligned} S(\beta) &= \sum_{i=1}^{K} [(y_i - X_i\beta)'(y_i - X_i\beta)] \\ &= \sum_{i=1}^{K} [y_i' y_i - 2y_i' X_i\beta + \beta' X_i' X_i \beta] \end{aligned} \tag{2.8}$$

for all individuals. Now by equating the derivatives of $S(\beta)$ with respect to β to 0, that is,

$$\frac{\partial S}{\partial \beta} = -2\sum_{i=1}^{K} [X_i' y_i - X_i' X_i \beta] = 0, \tag{2.9}$$

one obtains the OLS estimator of β as

$$\hat{\beta}_{OLS} = \left[\sum_{i=1}^{K} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' y_i\right], \tag{2.10}$$

which is the same as the moment estimator $\hat{\beta}_M$ of β given by (2.5). Consequently, $\hat{\beta}_{OLS}$ is unbiased for β with its covariance matrix as $V_{OLS} = V_M$ given by (2.6). Furthermore, asymptotically $(K \to \infty)$, V_{OLS} may be consistently estimated as $\hat{V}_{OLS} = \hat{V}_M$ by (2.7).

2.1.2.1 Generalized Least Squares (GLS) Method

In this approach, one takes the correlations of the data into account and minimizes the so-called generalized sum of squares

$$\begin{aligned} S^*(\beta) &= \sum_{i=1}^{K} [(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta)] \\ &= \sum_{i=1}^{K} [y_i'\Sigma_i^{-1} y_i - 2y_i'\Sigma_i^{-1} X_i\beta + \beta' X_i'\Sigma_i^{-1} X_i\beta] \end{aligned} \tag{2.11}$$

to obtain the GLS estimator of β. More specifically, equating the derivatives of $S^*(\beta)$ with respect to β to 0, that is,

$$\frac{\partial S^*}{\partial \beta} = -2\sum_{i=1}^{K} [X_i'\Sigma_i^{-1} y_i - X_i'\Sigma_i^{-1} X_i\beta] = 0, \tag{2.12}$$

one obtains the GLS estimator of β as

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{K} X_i'\Sigma_i^{-1} X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i'\Sigma_i^{-1} y_i\right]. \tag{2.13}$$

Because $E[Y_i] = X_i\beta$, it follows from (2.13) that $E[\hat{\beta}_{GLS}] = \beta$. Thus, $\hat{\beta}_{GLS}$ is an unbiased estimator of β, with its covariance given by

$$\begin{aligned} \text{cov}[\hat{\beta}_{GLS}] &= V_{GLS} \\ &= \left[\sum_{i=1}^{K} X_i'\Sigma_i^{-1} X_i\right]^{-1}, \end{aligned} \tag{2.14}$$

which, for unknown $\Sigma_i = \Sigma$, may be consistently estimated by

$$\hat{V}_{GLS} = \text{limit}_{K\to\infty} \left[\sum_{i=1}^{K} X_i' \hat{\Sigma}^{-1} X_i\right]^{-1}, \tag{2.15}$$

with $\hat{\Sigma} = K^{-1}\sum_{i=1}^{K}[(y_i - X_i\hat{\beta}_{GLS})(y_i - X_i\hat{\beta}_{GLS})']$. Note that if $\Sigma_i \neq \Sigma$ for $i = 1, \ldots, K$, the consistent estimation of Σ_i by using only T_i responses for the ith individual may or may not be easy. For example, if Σ_i is defined through a small number of common scale and/or correlation parameters those can be consistently estimated by using all $\{y_i, X_i\}$ for $i = 1, \ldots, K$; one may then easily obtain its consistent estimator. In other situations, the consistent estimation for Σ_i may not be so easy.

2.1.3 OLS Versus GLS Estimation Performance

Because both $\hat{\beta}_{OLS}$ (2.10) and $\hat{\beta}_{GLS}$ (2.13) are unbiased for β, they are consistent estimators. It follows, however, from (2.6) and (2.14) that their covariance matrices are not the same. Thus the variances of the two estimators given in the leading diagonals of the respective covariance matrices are likely to be different. Furthermore, it is known by the following theorem [see also Amemiya (1985, Section 6.1.3) and Rao (1973, Section 4a.2)] that the variances of the components of GLS estimator $\hat{\beta}_{GLS}$ are always smaller than the variances of the corresponding components of the OLS estimator $\hat{\beta}_{OLS}$. This makes $\hat{\beta}_{GLS}$ a more efficient estimator than the OLS estimator $\hat{\beta}_{OLS}$.

Theorem 2.1 For $u = 1, \ldots, p$, let $\hat{\beta}_{u,OLS}$ and $\hat{\beta}_{u,GLS}$ be the uth element of the OLS estimator $\hat{\beta}_{OLS}$ (2.10) and the GLS estimator $\hat{\beta}_{GLS}$, respectively. It then follows that

$$\text{var}[\hat{\beta}_{u,GLS}] \leq \text{var}[\hat{\beta}_{u,OLS}], \tag{2.16}$$

for all $u = 1, \ldots, p$, where 'var$[\cdot]$' represents the variance of the estimator in the square bracket.

Proof: Let $P_i = \Sigma_i^{-1}X_i$, $A = \left[\sum_{i=1}^{K} X_i'X_i\right]^{-1}$, and $B = \left[\sum_{i=1}^{K} X_i'\Sigma_i^{-1}X_i\right]^{-1}$. Then by (2.10) and (2.13), write

$$\begin{aligned}
\text{cov}[\hat{\beta}_{OLS}] &= \text{cov}\left[A\left(\sum_{i=1}^{K} X_i'Y_i\right) - B\left(\sum_{i=1}^{K} P_i'Y_i\right) + B\left(\sum_{i=1}^{K} P_i'Y_i\right)\right] \\
&= \text{cov}\left[A\left(\sum_{i=1}^{K} X_i'Y_i\right) - B\left(\sum_{i=1}^{K} P_i'Y_i\right)\right] + \text{cov}\left[\hat{\beta}_{GLS}\right],
\end{aligned} \tag{2.17}$$

by using the fact that

$$\text{cov}\left[\left\{A\left(\sum_{i=1}^{K} X_i'Y_i\right) - B\left(\sum_{i=1}^{K} P_i'Y_i\right)\right\}, \left\{B\left(\sum_{i=1}^{K} P_i'Y_i\right)\right\}\right] = 0. \tag{2.18}$$

It then follows from (2.17) that $\mathrm{var}[\hat{\beta}_{u,OLS}] \geq \mathrm{var}[\hat{\beta}_{u,GLS}]$, as in the theorem. We still need to show that (2.18) holds. We examine this directly as follows. Because $\mathrm{cov}(Y_i) = \Sigma_i$, and because all individuals are independent in the longitudinal setup, that is, $\mathrm{cov}(Y_i, Y_j) = 0$ for all $i \neq j,\ i, j = 1, \ldots, K$, we can write

$$\begin{aligned} &\mathrm{cov}\left[\left\{A\left(\sum_{i=1}^{K} X_i' Y_i\right) - B\left(\sum_{i=1}^{K} P_i' Y_i\right)\right\}, \left\{B\left(\sum_{i=1}^{K} P_i' Y_i\right)\right\}\right] \\ &= A\left(\sum_{i=1}^{K} X_i' \Sigma_i P_i\right) B' - B\left(\sum_{i=1}^{K} P_i' \Sigma_i P_i\right) B' \\ &= AA^{-1}B' - BB^{-1}B' = 0, \end{aligned} \tag{2.19}$$

by using $P_i = \Sigma_i^{-1} X_i$.

2.2 Estimation of β Under Stationary General Autocorrelation Structure

2.2.1 A Class of Autocorrelations

Recall from (2.2) that the $T_i \times T_i$ longitudinal covariance matrix for the ith individual is given by

$$\Sigma_i = A_i^{1/2} C_i A_i^{1/2},$$

where C_i is a $T_i \times T_i$ unknown correlation matrix. For convenience, one may express this correlation matrix as

$$C_i = (\rho_{i,ut}), \quad u, t = 1, \ldots, T_i, \tag{2.20}$$

with $\rho_{i,tt} = 1.0$. Note that in the linear longitudinal model setup, it is reasonable to assume that $\rho_{i,ut} = \rho_{ut}$ for all individuals $i = 1, \ldots, K$. The correlation matrix (2.20) may then be expressed as

$$C_i = (\rho_{ut}), \quad u, t = 1, \ldots, T_i, \tag{2.21}$$

which is a submatrix of a larger $T \times T$ correlation matrix

$$C = (\rho_{ut}), \quad u, t = 1, \ldots, T, \tag{2.22}$$

where $T = \max_{1 \leq i \leq K} T_i$. Note that once the C matrix is computed, C_i can be copied from C based on its dimension.

Further note that in the longitudinal set up, it is also quite reasonable to assume that the repeated responses follow a dynamic dependence model such as autoregressive moving average of order (q,r)(ARMA(q,r)) [Box and Jenkins (1976, Chapter 3)]. We note that ARMA(q,r) is a large class of autocorrelation structures used in general to explain the time effects in time series as well as in spatial data, among others. Under this large class of autocorrelations, the correlation structure in (2.21) may be expressed as

$$C_i(\rho) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T_i-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T_i-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T_i-1} & \rho_{T_i-2} & \rho_{T_i-3} & \cdots & 1 \end{bmatrix}, \tag{2.23}$$

where for $\ell = 1, \ldots, T_i$, ρ_ℓ is known to be the ℓth lag autocorrelation. Note that if the ARMA model is known for the repeated data, then these lag correlations in (2.23) may easily be computed. To understand this, consider the following examples.

Example 1: Autoregressive Order 1 (AR(1)) Structure

For $t = 1, \ldots, T_i$, re-write the tth equation for the ith individual from (2.1) as

$$y_{it} = x'_{it}\beta + \varepsilon_{it}, \tag{2.24}$$

and assume that

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + a_{it}, \tag{2.25}$$

with $|\rho| < 1$ and $a_{it} \overset{\text{iid}}{\sim} (0, \sigma_a^2)$. For a suitable integer r, one may exploit the recursive relation (2.25) and re-express ε_{it} as

$$\varepsilon_{it} = \rho^r \varepsilon_{i,t-r} + \sum_{j=0}^{r-1} \rho^j a_{i,t-j}. \tag{2.26}$$

Note that when the errors are assumed to be stationary, the joint distribution of

$$\varepsilon_{i,1-r}, \ldots, \varepsilon_{i,t-r}, \ldots, \varepsilon_{i,T_i-r}$$

remains the same for any $r = 0, \pm 1, \pm 2, \ldots, \pm\infty$. This is known as a strong stationarity condition. This strong condition is, however, not needed to find the stationary covariance matrix of the error vector ε_i. The relationship in (2.25) holds for any t in the stationary case, thus (2.26) may be written as

$$\varepsilon_{it} = \sum_{j=0}^{\infty} \rho^j a_{i,t-j}. \tag{2.27}$$

It then follows that

$$E[\varepsilon_{it}] = 0 \text{ and } \text{var}[\varepsilon_{it}] = \frac{\sigma_a^2}{1-\rho^2}, \tag{2.28}$$

for any $t = 1, \ldots, T_i$. Similarly, for $u < t = 2, \ldots, T_i$, by using the relationships

$$\varepsilon_{iu} = \sum_{j=0}^{\infty} \rho^j a_{i,u-j} \text{ and } \varepsilon_{it} = \sum_{j=0}^{t-u-1} \rho^j a_{i,t-j} + \rho^{t-u}[\sum_{j=0}^{\infty} \rho^j a_{i,u-j}], \tag{2.29}$$

one obtains the stationary covariance between ε_{iu} and ε_{it} as

$$\text{cov}[\varepsilon_{iu},\, \varepsilon_{it}] = \sigma_a^2 \frac{\rho^{t-u}}{1-\rho^2}. \tag{2.30}$$

It then follows from (2.28) and (2.30) that when the repeated responses

$$y_{i1}, \ldots, y_{it}, \ldots, y_{iT_i}$$

follow the AR(1) model (2.24)–(2.25), their means and variances are given by

$$E[Y_{it}] = x_{it}'\beta, \quad \text{var}[Y_{it}] = \sigma_a^2[1-\rho^2]^{-1}, \tag{2.31}$$

and their lag $|t-u|$ correlation $\rho_{|t-u|}$ (say) has the formula

$$\rho_{|t-u|} = \text{corr}[Y_{iu}, Y_{it}] = \rho^{|t-u|}, \text{ for } u \neq t,\ u, t = 1, \ldots, T_i, \tag{2.32}$$

where ρ is the model (2.25) parameter or may be referred to as the correlation index parameter. Here $|\rho| < 1$.

Note that the correlations in (2.32) satisfy the autocorrelation structure (2.23). Now, if the data were known to follow the AR(1) correlation model (2.24) – (2.25), one would then estimate the correlation structure in (2.23) by simply estimating $\rho_1 = \rho$ as this parameter determines all lag correlations as shown in (2.32). However, it may not be practical to assume that the data follow a specific structure such as AR(1), MA(1), or equicorrelation. Thus for more generality, we assume that the longitudinal data follow a general correlation structure (2.23) and estimate all lag correlations consistently by a suitable method of estimation. This is discussed in Section 2.2.

Example 2: Moving Average Order 1 (MA(1)) Structure

Suppose that as opposed to (2.25), the ε_{it} in (2.24) follows the model

$$\varepsilon_{it} = \rho a_{i,t-1} + a_{it}, \tag{2.33}$$

where ρ is a suitable scale parameter that does not necessarily have to satisfy $|\rho| < 1$, and a_{it} are white noise as in (2.25), that is, $a_{it} \overset{\text{iid}}{\sim} (0, \sigma_a^2)$. It is clear from (2.24) and (2.33) that the mean and the variance of y_{it} for all $t = 1,\ldots,T_i$ have the formulas

$$E[Y_{it}] = x'_{it}\beta, \quad \text{var}[Y_{it}] = \sigma_a^2(1+\rho^2), \tag{2.34}$$

and the lag $|t-u|$ correlations of the repeated responses have the formulas

$$\rho_{|t-u|} = \text{corr}(Y_{iu}, Y_{it}) = \begin{cases} \rho/(1+\rho) & \text{for } |t-u| = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.35}$$

The correlations in (2.35) also satisfy the autocorrelation structure (2.23).

Note that similar to the AR(1) and MA(1) models, the lag correlations for any higher order ARMA models such as ARMA(1,1) and ARMA(3,2) will also satisfy the autocorrelation structure (2.23). For the purpose of estimation, even if the data follow the MA(1) structure, we do not estimate the correlation structure by estimating the ρ in (2.35), rather, we estimate the general autocorrelation structure (2.23) which accommodates the correlation structure (2.35) as a special case.

Further note that there may be other correlation models yielding the autocorrelations as in (2.23). Consider the following model as an example.

Example 3: Equi-correlations (EQC) Structure

As a special case of the MA(1) model (2.33), we write

$$\varepsilon_{it} = \rho a_{i0} + a_{it}, \quad t = 1,\ldots,T_i, \tag{2.36}$$

where a_{i0} is considered to be an error value occurred at an initial time, and ρ is a suitable scale parameter. Assume that

$$a_{it} \overset{\text{iid}}{\sim} (0, \sigma_a^2), \text{and also} \quad a_{i0} \sim (0, \sigma_a^2),$$

and a_{it} and a_{i0} are independent for all t. It then follows from (2.24) and (2.36) that the mean and the variance of y_{it} are given by

$$E[Y_{it}] = x'_{it}\beta, \quad \text{var}[Y_{it}] = \sigma_a^2(1+\rho^2),$$

as in (2.34), but the lag correlations have the formulas

$$\rho_{|t-u|} = \text{corr}(Y_{iu}, Y_{it}) = \rho^2/(1+\rho^2), \tag{2.37}$$

for all lags $|t-u| = 1,\ldots,T_i - 1$. This equicorrelation structure (2.37) is also accommodated by the general autocorrelation structure (2.23).

2.2.2 Estimation of β

The $\hat{\beta}_{GLS}$ in (2.13) is the best among linear unbiased estimators for β, therefore we may still use the formula

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{K} X_i' \Sigma_i^{-1} X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' \Sigma_i^{-1} y_i\right], \tag{2.38}$$

but under the current special autocorrelation class, we estimate Σ_i as

$$\hat{\Sigma}_i = A_i^{1/2} C_i(\hat{\rho}) A_i^{1/2}, \tag{2.39}$$

where the $C_i(\hat{\rho})$ matrix is computed by (2.23) by replacing ρ_ℓ with an approximate unbiased moment estimator $\hat{\rho}_\ell$ (say).

Now to compute the $C_i(\hat{\rho})$ matrix in (2.39), in light of (2.22), we first compute the larger $C(\hat{\rho})$ matrix for $\ell = 1, \ldots, T-1$, where $T = \max_{1 \le i \le K} T_i$ for $T_i \ge 2$. Suppose that δ_{it} is an indicator variable such that

$$\delta_{it} = \begin{cases} 1 & \text{if } t \le T_i \\ 0 & \text{if } T_i < t \le T. \end{cases}$$

for all $t = 1, \ldots, T$. For known β and σ_{itt}, the ℓth lag correlation estimate $\hat{\rho}_\ell$ for the larger $C(\hat{\rho})$ matrix may be computed as

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \delta_{it} \delta_{i,t+\ell} [\left(\frac{y_{it} - x_{it}'\beta}{\sigma_{itt}}\right) \left(\frac{y_{i,t+\ell} - x_{it,t+\ell}'\beta}{\sigma_{i,t+\ell,t+\ell}}\right)] / \sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \delta_{it} \delta_{i,t+\ell}}{\sum_{i=1}^{K} \sum_{t=1}^{T} \delta_{it} [\frac{y_{it} - x_{it}'\beta}{\sigma_{itt}}]^2 / \sum_{i=1}^{K} \delta_{it}}, \tag{2.40}$$

[cf. Sneddon and Sutradhar (2004, eqn. (16)) in a more general linear longitudinal setup] for $\ell = 1, \ldots, T-1$. Note that as this estimator contains $\hat{\beta}_{GLS}$, both (2.38) and (2.40) have to be computed iteratively until convergence.

Further note that $\hat{\rho}_\ell$ in (2.40) is an approximately unbiased estimator of ρ_ℓ. This is because irrespective of the autocorrelation structure for the repeated data, it follows that

$$\begin{aligned} E[\hat{\rho}_\ell] &\simeq \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \delta_{it} \delta_{i,t+\ell} E[\left(\frac{y_{it} - x_{it}'\beta}{\sigma_{itt}}\right) \left(\frac{y_{i,t+\ell} - x_{it,t+\ell}'\beta}{\sigma_{i,t+\ell,t+\ell}}\right)] / \sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \delta_{it} \delta_{i,t+\ell}}{\sum_{i=1}^{K} \sum_{t=1}^{T} \delta_{it} E[\frac{y_{it} - x_{it}'\beta}{\sigma_{itt}}]^2 / \sum_{i=1}^{K} \delta_{it}} \\ &= \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \delta_{it} \delta_{i,t+\ell} [\rho_\ell] / \sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \delta_{it} \delta_{i,t+\ell}}{\sum_{i=1}^{K} \sum_{t=1}^{T} \delta_{it} / \sum_{i=1}^{K} \delta_{it}} \\ &= \rho_\ell \end{aligned} \tag{2.41}$$

It then also follows that $\hat{\rho}_\ell$ in (2.40) is a consistent $(K \to \infty)$ estimator for ρ_ℓ and its use in (2.38) does not alter the efficiency property of $\hat{\beta}_{GLS}$ when computed assuming that ρ is known. In practice, $\hat{\beta}_{GLS}$ from (2.38) is used for β in (2.40). Furthermore, in a linear model, it is likely that σ_{itt} are independent of i and may be written as $\sigma_t^2 \equiv \sigma_{itt}$ for all $i = 1,\ldots,K$. Now for the estimation of σ_t^2, or in general, for the estimation of the $A_i =$ diagonal $[\sigma_1^2,\ldots,\sigma_{T_i}^2]$ in (2.39), we may obtain the estimate of σ_t^2 for all $t = 1,\ldots,T$, by the method of moments using the formula

$$\hat{\sigma}_t^2 = \sum_{i=1}^{K} \delta_i [y_{it} - x_{it}' \hat{\beta}_{GLS}]^2 / \sum_{i=1}^{K} \delta_i, \tag{2.42}$$

where

$$\delta_i = \begin{cases} 1 & \text{if } \delta_{ij} = 1 \text{ for all } 1 \le j \le t \\ 0 & \text{otherwise,} \end{cases}$$

with δ_{ij} defined as in (2.40).

Note that the computation of the inverse matrix Σ_i^{-1} in (2.38) requires the inversion of the general lag correlation matrix $C_i = (\rho_{|u-t|})$. This may be easily done by using any standard software such as *FORTRAN-90, R,* or *S-PLUS*. For specific AR(1) (2.32), MA(1) (2.35), and EQC (2.37) structures, C_i^{-1} may, however, be calculated directly by using the formulas given in Exercises 5, 6, and 7, respectively.

2.3 A Rat Data Example

As an illustration for the application of the linear longitudinal fixed model (LLFM) described through $(2.1) - (2.2)$ with general autocorrelation matrix $C_i(\rho)$ as in (2.23), we consider the biological longitudinal experimental data, originally obtained by the Department of Nutrition, University of Guelph, and subsequently analyzed by other researchers such as Srivastava and Carter (1983, pp. 146 – 150). For convenience we reproduce this data as shown in Tables 2A and 2B in the Appendix. This dataset contains the longitudinal food habits of 32 rats over a period of six days under two different situations. First, for six days all 32 rats were given a control diet. Next, these 32 rats were divided equally into four groups and four different treatment diets (containing four different amounts of phosphorous) were given, and the amount of food eaten by eight rats in each group was recorded over another six days. As far as the covariates are concerned, the initial weight for each of the 32 rats was recorded and it was of interest to see the effect of these initial weights on food habits for six days. We give some summary statistics for these data in Table 2.1 below.

Note that to understand the effect of initial weight on the longitudinal food habits, one has more information here for the control group as compared to any of the individual treatment groups. This is because all 32 rats were given the control diet

Table 2.1 Summary statistics for food amount eaten by the rats under the control and treatment diets.

		Day					
Group	Statistic	1	2	3	4	5	6
Control (0.1% P)	Average amount	11.19	10.50	8.17	7.95	7.93	8.46
	Standard deviation	2.97	4.25	3.61	3.35	3.72	3.73
TrG1 (0.25% P)	Average amount	6.93	6.84	5.72	9.26	8.65	8.28
	Standard deviation	4.01	2.68	3.56	2.90	2.20	2.36
TrG2 (0.65% P)	Average amount	6.89	9.69	8.92	9.70	10.88	9.52
	Standard deviation	3.33	2.00	3.18	3.57	3.81	2.40
TrG3 (1.3% P)	Average amount	7.56	8.89	6.40	6.05	6.46	7.70
	Standard deviation	2.91	5.42	4.79	3.04	3.40	3.71
TrG4 (1.71% P)	Average amount	6.54	5.49	4.11	4.54	5.73	3.66
	Standard deviation	3.00	4.10	2.17	2.28	2.35	1.89

based food for six days, and each treatment group had 8 rats to feed over six days. Under the circumstances, it is appropriate to fit two linear longitudinal models, one for the control group and the other for the treatment groups.

For the control group, following (2.1), we fit the model

$$y_{it} = \beta_{c,0} + x_{i,INW}\beta_{c,1} + \varepsilon_{it}, \quad \text{for } t = 1,\ldots,6;\ i = 1,\ldots,32, \tag{2.43}$$

where y_{it} is the amount of control diet based food eaten by the ith rat on the tth day, $x_{i,INW}$ denote the initial weight of the ith rat which is independent of time, and ε_{it} is the corresponding error. Note that for convenience, we have defined the initial weight $x_{i,INW}$ as a standardized quantity. That is,

$$x_{i,INW} = \frac{TIW_i - MIW}{STDIW} = \frac{TIW_i - 290.25}{6.98},$$

where TIW_i is the true initial weight of the ith $(i = 1,\ldots,32)$ rat, MIW and $STDIW$ are the mean and the standard deviation of the initial weights of the 32 rats. Furthermore, in (2.43), $\beta_{c,0}$ and $\beta_{c,1}$ denote the regression effects under the control group. Because the food eaten by the same rat over $T = 6$ days must be correlated, following (2.23) we assume that $\varepsilon_{i1},\ldots,\varepsilon_{iT}$ follow an autocorrelation class with $T \times T$ constant correlation matrix for all $i = 1,\ldots,32$, given by

$$C_i(\rho) \equiv C = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix}, \tag{2.44}$$

ρ_ℓ being the ℓth lag autocorrelation, for $\ell = 1,\ldots,T-1$. For the control group, the moment estimates for the lag correlations were found to be

$$\hat{\rho}_1 = 0.55, \quad \hat{\rho}_2 = 0.31, \quad \hat{\rho}_3 = 0.22, \quad \hat{\rho}_4 = 0.17, \; \hat{\rho}_5 = -0.01,$$

and the GLS estimate of $\beta_{c,0}$ and $\beta_{c,1}$ with their standard errors (s.e.) were found to be

$$\hat{\beta}_{c,0} = 9.34, \quad \hat{\beta}_{c,1} = 0.40,$$

and

$$\text{s.e.}(\hat{\beta}_{c,0}) = 0.42, \quad \text{s.e.}(\hat{\beta}_{c,1}) = 0.42,$$

respectively. The estimates for the lag correlations show an exponential decay. As expected, the correlations tend to decrease as the lag increases. Thus, the food amount eaten on day 3, for example, is more highly correlated with the day 2 amount as compared to the day 1 amount. This explains the nature of the time effects on the food habits of the rats when they are given control diet based food.

Note that to compute $\hat{\beta}_{GLS}$ by (2.38) and $\hat{\rho}_\ell$ by (2.40), we have used $\sigma_{itt} = \sigma_t^2$, which in turn was estimated by (2.42). For the control group data, these estimates for $t = 1, \ldots, 6$, were found to be

$$\hat{\sigma}_1^2 = 12.01, \quad \hat{\sigma}_2^2 = 18.84, \quad \hat{\sigma}_3^2 = 14.13, \quad \hat{\sigma}_4^2 = 13.37, \quad \hat{\sigma}_5^2 = 15.89, \quad \hat{\sigma}_6^2 = 14.39.$$

We now interpret the effect of the initial weight of a rat on the food habit under the control group. The initial weight has a regression effect of 0.40 on the amount of food eaten by a rat. This value along with the intercept estimate 9.34 indicates that a rat with initial weight between 276.29 and 304.21 units, for example, has eaten at a given day an amount of food that ranges between $9.34 - 2 \times 0.40 = 8.54$ and $9.34 + 2 \times 0.40 = 10.14$ units. Note that under the control group, the first row in the summary statistics in Table 2.1 shows that a rat on the average has eaten food ranging from 7.93 to 10.50 units over five days with an exception of 11.19 units of food eaten on the first day. Thus, in general the estimated food amount yielded by the model (2.43) – (2.44) appears to agree with the summary statistics under the control group.

In order to write a linear longitudinal model for the treatment group, we first consider three indicator covariates to represent four treatment groups. For $i = 1, \ldots, 32$, let $x_{i1,Tr}$, $x_{i2,Tr}$ and $x_{i3,Tr}$ be the three indicator covariates such that $x_{i1,Tr} = 0$, $x_{i2,Tr} = 0$, $x_{i3,Tr} = 0$ indicate that the ith individual is assigned to treatment group 1 (TrG1). Similarly, the ith individual rat belongs to

$$\begin{aligned} &\text{TrG2 when } x_{i1,Tr} = 1, \quad x_{i2,Tr} = 0, \quad x_{i3,Tr} = 0; \ \text{ or,} \\ &\text{TrG3 when } x_{i1,Tr} = 0, \quad x_{i2,Tr} = 1, \quad x_{i3,Tr} = 0; \ \text{ or,} \\ &\text{TrG4 when } x_{i1,Tr} = 0, \quad x_{i2,Tr} = 0, \quad x_{i3,Tr} = 1. \end{aligned}$$

Now, the model under the treatment group, as opposed to (2.43) for the control group, may be written as

$$y_{it} = \beta_{Tr,0} + x_{i,INW}\beta_{Tr,1} + x_{i1,Tr}\beta_{Tr,2} + x_{i2,Tr}\beta_{Tr,3}$$

$$+x_{i3,Tr}\beta_{Tr,4} + \varepsilon_{it}, \text{ for } t = 1,\ldots,8;\ i = 1,\ldots,32, \qquad (2.45)$$

We now apply the model (2.45) to the rat data in Tables 2A and 2B in the appendix and obtain the regression effects including the treatment group effects by using the formulas (2.38) and (2.39) with $C_i(\rho) = C$ as in (2.44), and $A_i =$ diagonal $[\sigma_1^2,\ldots,\sigma_6^2]$. The lag correlations necessary to compute the regression effects were estimated by using the moment estimating equation (2.40). These estimates for the lag correlations are:

$$\hat{\rho}_1 = 0.39,\quad \hat{\rho}_2 = 0.14,\quad \hat{\rho}_3 = 0.27,\quad \hat{\rho}_4 = 0.05,\quad \hat{\rho}_5 = -0.18.$$

Note that as compared to the control group, the lag correlations are relatively smaller in the treatment group. Also, unlike the control group, there appears to be a spike for the lag 3 correlations even though there is a general tendency of decay in correlations as lag increases. Thus, the time effects in the control and treatment groups appear to be generally different on the food habits of the rats.

The GLS estimates of the regression effects including the treatment group effects and their standard errors were found to be

$$\hat{\beta}_{Tr,0} = 8.05,\ \hat{\beta}_{Tr,1} = 0.72,\quad \hat{\beta}_{Tr,2} = 0.95,\quad \hat{\beta}_{Tr,3} = -0.89,\quad \hat{\beta}_{Tr,4} = -3.12,$$

and

$$\text{s.e.}(\hat{\beta}_{Tr,0}) = 0.63,\quad \text{s.e.}(\hat{\beta}_{Tr,1}) = 0.32,\quad \text{s.e.}(\hat{\beta}_{Tr,2}) = 0.89,$$
$$\text{s.e.}(\hat{\beta}_{Tr,3}) = 0.91,\quad \text{s.e.}(\hat{\beta}_{Tr,4}) = 0.90,$$

respectively. Note that under the treatment group, the initial weight has a larger regression effect of 0.72 on the amount of food eaten by a rat, as compared to 0.40 in the control group. Because $x_{i1,Tr} = 0$, $x_{i2,Tr} = 0$, $x_{i3,Tr} = 0$, for the treatment group 1 (TrG1), the initial weight effect 0.72 along with the intercept estimate 8.04 indicates that a rat in the TrG1 with initial weight between 276.29 and 304.21 units, for example, has eaten at a given day an amount of food ranging between $8.05 - 2 \times 0.72 = 6.61$ and $8.05 + 2 \times 0.72 = 9.49$ units. These estimated food amounts are smaller than the estimated food amounts found under the control group. The food amount eaten by the rats under the TrG1 in row 3 of Table 2.1 are in general less than those under the control group shown in row 1, thus the linear longitudinal models (2.43) and (2.45) appear to explain the data well for the control and treatment groups, respectively. Further note that Table 2.1 shows that the amount of food eaten by the rats under the TrG2 (row 5) over the six days are in general larger than those eaten by the rats in TrG1 (row 3). But, the amount of food eaten by the rats under the TrG3 (row 7) and TrG4 (row 9) over the six days tends to be smaller than that eaten by the rats in TrG1 (row 3). The positive value of the TrG2 effect $\hat{\beta}_{Tr,2} = 0.95$, and the negative values of the TrG3 and TrG4 effects, that is, $\hat{\beta}_{Tr,3} = -0.89$, and $\hat{\beta}_{Tr,4} = -3.12$, respectively, fully support the longitudinal food habits of the rats in TrG2, TrG3, and TrG4, as compared to those in TrG1.

The estimates for the variance components for the treatment group were found to be

$$\hat{\sigma}_1^2 = 12.35, \quad \hat{\sigma}_2^2 = 16.00, \quad \hat{\sigma}_3^2 = 14.26, \quad \hat{\sigma}_4^2 = 8.90, \quad \hat{\sigma}_5^2 = 10.20, \quad \hat{\sigma}_6^2 = 7.38.$$

2.4 Alternative Modelling for Time Effects

Note that in the last section time effects on the repeated responses are explained through the lag correlations of these responses. Some authors, for example, Pearson et al. (1994), Verbeke and Molenberghs (2000, Chapter 3), and Verbeke and Lesaffre (1999), modelled the repeated responses in a mixed model setup as a linear or quadratic function over time. In the present fixed model (2.1) set up, these models may be expressed as

$$y_{it} = [x_i'\alpha]t + [x_i'\beta]t^2 + \varepsilon_{it}, \tag{2.46}$$

[cf. Verbeke and Molenberghs (2000, Chapter 3, eqn. (3.5))] where x_i is the p-dimensional time-independent covariate vector, and α and β are the effects of tx_i and t^2x_i on the response y_{it}, and $\varepsilon_{it} \overset{\text{iid}}{\sim} (0, \sigma^2)$. It is clear from (2.46) that time is considered here as a deterministic factor and hence one is unable to model the correlations among the repeated responses. Diggle, Liang, and Zeger (1994) [see also Verbeke and Molenberghs (2000, Chapter 3, eqn. (3.5))] argue that the effect of serial (lag) correlations is very often dominated by suitable random effects and consequently model the correlations of the repeated data through the introduction of random effects. This may be done by modifying the model in (2.46) as

$$y_{it} = [x_i'\alpha + \gamma_{i1}]t + [x_i'\beta + \gamma_{i2}]t^2 + \varepsilon_{it}, \tag{2.47}$$

[cf. Verbeke and Molenberghs (2000, Chapter 3, eqn. (3.10))] or as

$$y_{it} = x_{it}'\beta + z_{i1}\gamma_{i1} + z_{i2}\gamma_{i2} + \varepsilon_{it}, \tag{2.48}$$

[cf. Verbeke and Molenberghs (2000, Chapter 3, eqn. (3.11))] where z_{i1} and z_{i2} are suitable covariates, and the random effects γ_{i1} and γ_{i2} may be independent or correlated with marginal properties

$$\gamma_{i1} \overset{\text{iid}}{\sim} (0, \sigma_{\gamma_1}^2) \text{ and } \gamma_{i2} \overset{\text{iid}}{\sim} (0, \sigma_{\gamma_2}^2).$$

But, as follows from Sneddon and Sutradhar (2004), even though the random effects γ_{i1} and γ_{i2} in (2.47) and (2.48) generate an equicorrelation structure for the repeated responses, they do not appear to address the time effects. This is because these individual specific random effects remain the same throughout the data collection period and hence cannot represent any time effects. Nevertheless, the mixed model (2.48) is interesting in its own right and we discuss this model in the next chapter in a wider setup under the assumption that $\varepsilon_{i1},\ldots,\varepsilon_{it},\ldots,\varepsilon_{iT_i}$ follow a class of general autocorrelation structures as introduced in Section 2.2.1.

Exercises

2.1. (Section 2.1.2) [Best linear unbiased estimator]
Consider the model $y_i = X_i\beta + \varepsilon_i$ (2.1) but with the assumption that $\varepsilon_i \stackrel{\text{iid}}{\sim} (0, \sigma^2 I_{T_i})$. Now consider all linear unbiased estimators of β in the form $\beta^* = \sum_{i=1}^{K} Q_i' y_i$ satisfying $\sum_{i=1}^{K} Q_i' X_i = I_p$, with Q_i as the $T_i \times p$ constant matrix and I_p as the $p \times p$ identity matrix. Show that $\hat{\beta}_{OLS} = \left[\sum_{i=1}^{K} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' y_i\right]$ in (2.10) belongs to this class and is better than β^*; that is $\text{var}[\hat{\beta}_{OLS}] \leq \text{var}[\beta^*]$.

2.2. (Section 2.1.3)
Similar to that of Exercise 2.1, argue that $\hat{\beta}_{GLS} = \left[\sum_{i=1}^{K} X_i' \Sigma_i^{-1} X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' \Sigma_i^{-1} y_i\right]$ in (2.13) also belongs to the class of linear unbiased estimators and show that $\hat{\beta}_{GLS}$ is the best linear unbiased estimator in this class for correlated data satisfying the assumption that $\varepsilon_i \stackrel{\text{id}}{\sim} (0, \Sigma_i)$ as in model (2.1) instead of $\varepsilon_i \stackrel{\text{iid}}{\sim} (0, \sigma^2 I_{T_i})$ as imposed in Exercise 2.1.

2.3. (Section 2.1.4) [An alternative indirect proof for Theorem 2.1]
Suppose that the data following the model (2.1) are correlated. It then follows from Exercise 2.2 that $\hat{\beta}_{GLS}$ given by (2.13) is the best linear unbiased estimator of β. Use this result and argue that $\hat{\beta}_{GLS}$ is better than the independence assumption based OLS estimator $\hat{\beta}_{OLS}$ (2.10).

2.4. (Section 2.2.1) [Alternative proofs under the AR(1) process]
When the errors in the AR(1) process (2.24) are stationary, it follows that $E[\varepsilon_{it}^2] = E[\varepsilon_{i,t-1}^2] = \sigma^2$, for all t. Use this result and show by (2.24) that

$$\text{var}[\varepsilon_{it}] = \sigma^2 = \frac{\sigma_a^2}{1-\rho^2} \;\text{ and }\; \text{cov}[\varepsilon_{iu}, \varepsilon_{it}] = \sigma_a^2 \frac{\rho^{|t-u|}}{1-\rho^2}.$$

2.5. (Section 2.2.1) [Inversion of the AR(1) process based correlation matrix (2.32)]
The inversion of the AR(1) correlation matrix

$$C_i(\rho_{|t-u|}) = (\rho^{|t-u|}), \;\text{ for } u \neq t,\; u,t = 1,\ldots,T_i,$$

has the form

$$C_i^{-1}(\rho) = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho & 0 & 0 & \cdots & 0 & 0 \\ \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \\ 0 & 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix},$$

[Kendall, Stuart, and Ord (1983, p. 614)].

2.6. (Section 2.2.1) [Inversion of the MA(1) process based correlation matrix (2.35)] Suppose that for $\theta_1 = -\theta/(1+\theta^2)$, the $T_i \times T_i$ correlation matrix for the MA(1) process is written as

$$C_i(\theta) = \begin{bmatrix} 1 & \theta_1 & 0 & 0 & \cdots & 0 & 0 \\ \\ \theta_1 & 1 & \theta_1 & 0 & \cdots & 0 & 0 \\ 0 & \theta_1 & 1 & \theta_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & \theta_1 \\ 0 & 0 & 0 & 0 & \cdots & \theta_1 & 1 \end{bmatrix}.$$

For $u,t = 1,\ldots,T_i$, the $(u,t)th$ element of the $C_i^{-1}(\theta)$ matrix is given by

$$\frac{1+\theta^2}{1-\theta^2}\left[\left\{\theta^{|u-t|} - \theta^{2(T_i+2)-u-t-2}\right\}\right.$$
$$\left. -\frac{\theta^{u+t}}{1-\theta^{2(T_i+2)-2}}\left\{(1-\theta^{2(T_i+2)-2u-2})(1-\theta^{2(T_i+2)-2t-2})\right\}\right],$$

[Sutradhar and Kumar (2003, Section 2)]. The inverse of the $C_i(\rho)$ matrix in (2.35) may then easily be computed by using θ in terms of ρ derived from the relationship $-\theta/(1+\theta^2) = \rho/(1+\rho)$.

2.7. (Section 2.2.1) [Inversion of the EQC process based correlation matrix (2.37)] The inversion of the $T_i \times T_i$ EQC matrix

$$C_i(\theta) = (1-\theta)I_{T_i} + \theta U_{T_i}$$

with I_{T_i} and U_{T_i} as the identity and unit matrices, respectively, has the form given by

$$C_i^{-1}(\theta) = (a-b)I_{T_i} + bU_{T_i},$$

[Seber (1984, p. 520)] where

$$a = \frac{1+(T_i-2)\theta}{(1-\theta)\{1+(T_i-1)\theta\}} \text{ and } b = -\frac{\theta}{(1-\theta)\{1+(T_i-1)\theta\}}.$$

The inverse of the $C_i(\rho)$ matrix in (2.37) may then be computed by using $\theta = \rho^2/(1+\rho^2)$.

References

1. Amemiya, T. (1985). *Advanced Econometrics.* Cambridge, MA: Harvard University Press.
2. Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis Forecasting and Control.* San Francisco: Holden-Day.
3. Diggle, P. J., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford Science. Oxford: Clarendon Press.
4. Kendall, M., Stuart, A., & Ord, J. K. (1983). *The Advanced Theory of Statistics, Vol. 3,* London: Charles Griffin.
5. Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.
6. Pearson, J. D., Morrell, C. H., Landis, P. K., Carter, H. B., & Brant, L. J. (1994). Mixed-effects regression models for studying the natural history of prostate disease. *Statist. Med.,* 13, 587 – 601.
7. Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* New York: John Wiley & Sons.
8. Seber, G. A. F. (1984). *Multivariate Observations.* New York: John Wiley & Sons.
9. Sneddon, G. & Sutradhar, B. C. (2004). On semi-parametric familial-longitudinal models. *Statist. Probab. Lett.,* 69, 369 – 379.
10. Srivastava, M. S. & Carter, E. M. (1983). *An Introduction to Applied Multivariate Statistics.* New York: North-Holland.
11. Sutradhar, B. C. & Kumar, P. (2003). The inversion of the correlation matrix for MA(1) process. *Appl. Math. Lett.,* 16, 317 – 321.
12. Verbeke, G. & Lesaffre, E. (1999). The effect of drop-out on the efficiency of longitudinal experiments. *Appl. Statist.,* 48, 363 – 375.
13. Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York: Springer.

Appendix

Table 2A: Rat Data with Control Diet

Group	Initial Weight	Days 1	2	3	4	5	6
Control	254	12.7	10.4	5.1	8.6	7.1	9.7
	262	7.2	8.4	8.5	6.8	6.3	4.5
	301	14.8	13.9	8.4	7.3	8.0	10.4
	311	5.6	10.2	7.8	6.1	6.4	16.5
	290	13.9	12.1	8.8	8.8	8.1	7.8
	300	10.4	11.2	12.5	7.0	6.9	6.9
	306	16.6	17.8	14.0	6.8	5.9	5.3
	286	13.9	14.3	5.9	7.7	9.2	5.7
Control	275	11.9	7.0	5.9	6.1	0.8	5.1
	282	10.7	11.3	4.4	3.9	4.7	5.3
	256	10.1	6.9	7.8	6.4	9.5	7.9
	276	10.8	5.2	1.3	1.3	2.1	6.3
	337	14.7	14.4	11.6	7.4	7.8	14.8
	296	9.7	12.1	5.2	9.1	9.7	5.2
	309	5.5	7.1	7.8	3.1	1.5	8.4
	296	13.1	6.5	1.3	0.9	0.8	0.5
Control	275	8.8	17.7	11.5	6.6	5.4	12.0
	292	8.3	3.2	5.2	8.9	4.3	4.4
	338	16.2	11.9	10.2	15.6	15.3	13.9
	248	7.7	4.9	11.7	12.7	13.2	10.7
	315	14.5	14.0	16.9	8.4	13.1	9.8
	295	11.6	2.5	5.5	4.5	5.8	8.6
	312	5.3	6.1	1.5	4.1	6.2	2.1
	286	11.2	11.0	5.7	8.1	10.0	8.1
Control	275	13.5	9.7	12.3	13.4	14.0	6.1
	270	11.6	2.4	9.7	14.0	10.8	10.3
	290	10.0	14.8	9.1	9.6	8.2	9.3
	260	12.3	16.2	6.6	9.2	8.3	12.6
	302	13.6	14.9	9.3	10.2	11.5	15.8
	284	12.8	13.2	11.6	11.5	11.1	10.5
	280	10.9	14.3	10.8	9.6	13.2	10.0
	329	8.3	10.5	7.5	10.6	8.5	6.2

Table 2B: Rat Data with Treatment Diet

Group	Initial Weight	Days 1	2	3	4	5	6
TrG1	254	3.0	4.9	3.8	5.5	6.3	5.3
	262	7.9	7.0	7.7	8.7	7.1	11.4
	301	6.0	7.4	7.2	10.5	12.7	8.9
	311	16.7	12.8	11.6	15.9	10.6	4.5
	290	4.8	6.5	4.1	7.1	7.0	7.6
	300	6.0	7.3	1.2	9.0	10.8	8.1
	306	3.7	2.7	1.0	7.7	7.4	11.5
	286	7.3	6.1	9.2	9.7	7.3	9.0
TrG2	275	4.2	9.7	10.1	7.7	15.3	10.9
	282	6.6	8.1	9.9	8.8	11.4	10.1
	256	5.7	9.8	4.7	5.9	4.5	7.6
	276	6.6	8.9	12.9	15.0	13.0	8.6
	337	9.7	9.9	8.9	15.5	11.0	5.0
	296	8.8	6.5	7.3	5.7	6.0	12.2
	309	12.6	13.9	4.2	11.1	16.0	8.9
	296	0.9	10.7	13.4	7.9	9.8	12.9
TrG3	275	2.3	0.6	1.0	7.0	9.8	4.8
	292	8.1	11.8	7.3	7.0	10.3	9.5
	338	5.6	2.0	0.0	0.8	3.6	8.9
	248	9.0	7.5	0.7	1.4	0.4	0.3
	315	6.6	6.0	9.7	9.8	5.1	6.0
	295	6.0	16.4	9.9	8.8	9.4	8.4
	312	11.7	14.7	13.4	7.0	4.2	13.3
	286	11.2	12.1	9.2	6.6	8.9	10.4
TrG4	275	3.3	1.2	1.9	1.3	2.8	5.8
	270	5.7	15.5	3.8	1.8	6.3	1.5
	290	7.2	4.1	7.0	3.4	6.5	1.3
	260	8.1	4.8	7.9	8.2	9.9	5.2
	302	2.7	5.6	2.7	4.3	3.0	1.3
	284	6.2	6.4	2.5	4.1	4.9	4.0
	280	6.0	2.2	2.0	7.0	4.1	4.2
	329	13.1	4.1	5.1	6.2	8.3	6.0

Chapter 3
Overview of Linear Mixed Models for Longitudinal Data

Recall from the last chapter [eqn. (2.48)] that there exists [Verbeke and Molenberghs (2000, Chapter 3, eqn. (3.11)); Diggle, Liang, and Zeger (1994)] a random effects based longitudinal mixed model given by

$$y_{it} = x'_{it}\beta + z_i\gamma_i + \varepsilon_{it}, \tag{3.1}$$

where the ε_{it} are independent errors for all $t = 1,\ldots,T_i$ for the ith $(i = 1,\ldots,K)$ individual. This model (3.1) introduces the lag correlations through the random effects γ_i. For example, for

$$\gamma_i \overset{\text{iid}}{\sim} (0,\sigma_\gamma^2) \text{ and } \varepsilon_{it} \overset{\text{iid}}{\sim} (0,\sigma_\varepsilon^2) \tag{3.2}$$

and when it is assumed that γ_i and ε_{it} are independent, it may be shown that all lag correlations under the model $(3.1)-(3.2)$ are given by

$$\text{corr}(Y_{it}, Y_{it'}) = \rho_{|t-t'|} = \frac{z_i^2\sigma_\gamma^2}{\sigma_\varepsilon^2 + z_i^2\sigma_\gamma^2}, \tag{3.3}$$

yielding equal correlations between any two responses of the ith individual. Note that it is not only that the model (3.1) is limited to the equicorrelation structure, but these correlations also do not appear to accommodate the time effects in the longitudinal responses. This is because the random effect γ_i under the model (3.1) remains the same during the collection of the repeated data $y_{i1},\ldots,y_{iT_i}$, indicating that γ_i cannot represent the time effects.

Note, however, that there is a long history of using the random effects model (3.1) in the statistics and econometrics literature. See, for example, Searle (1971, Chapter 9) and the references therein. See also Amemiya (1985, Section 6.6.2). To be specific, the random effects model (3.1) is considered to be a variance component model in the linear model setup, and this is used mainly to analyze clustered or familial data such as (1) the independent responses collected from the members of the same family, and (2) the independent responses collected from a group of individuals exposed to the same treatment. As far as the inferences for the variance com-

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_3,

ponents of the random random effects model (3.1) are concerned, there exist many techniques such as (a) ANOVA (analysis of variance) or moment estimation [Searle (1971)], (b) quadratic estimator for the balanced ($T_i = T$) cases [LaMotte (1973); Mathew, Sinha and Sutradhar (1992)], and (c) non-quadratic estimation [Chow and Shao (1988); Sutradhar (1997)]. There also exists restricted maximum likelihood estimation [Herbach (1959); Thompson (1962)] for the nonnegative estimation of the variance components provided it is known that the random effects γ_i and the errors ε_{it} follow a known distribution such as the normal distribution.

Turning back to the introduction of the time effects in a linear mixed model, one may attempt to use the time-dependent random effects and rewrite the model (3.1) as

$$y_{it} = x_{it}'\beta + z_i \gamma_{it} + \varepsilon_{it}, \tag{3.4}$$

where $\gamma_{i1}, \ldots, \gamma_{iT_i}$ may be assumed to have a $T_i \times T_i$ suitable covariance structure. Note, however, that this model (3.4) encounters several technical difficulties. For example, for the case $z_i = 1$, $\gamma_{it} + \varepsilon_{it}$ may be considered as a new error and it may not be possible to identify the individual contribution of γ_{it} and ε_{it} to the variance of the data y_{it}. Furthermore, it is not practical to assume that the individual effect gets changed with respect to time especially when longitudinal data are collected for a short period from the same individual.

3.1 Linear Longitudinal Mixed Model

As opposed to the model (3.4), we now write a suitable linear mixed model in such a way that the individual random effect remains unchanged during the data collection period but the responses are still longitudinally correlated. This type of correlation model conditional on the random effects may be constructed by using a suitable autocorrelation structure for the error components ε_{it} in $(3.1)-(3.2)$ for $t = 1, \ldots, T_i$. For the purpose, we first re-express the model $(3.1)-(3.2)$ as

$$y_i = X_i\beta + 1_{T_i} z_i \gamma_i + \varepsilon_i, \tag{3.5}$$

where

$$y_i = [y_{i1}, \ldots, y_{iT_i}]', \quad X_i' = [x_{i1}, \ldots, x_{iT_i}], \quad \varepsilon_i = [\varepsilon_{i1}, \ldots, \varepsilon_{iT_i}]',$$

and 1_{T_i} is the T_i-dimensional unit vector. Note, however, that because in practice the covariates z_i associated with the random effects γ_i may not be available, it is customary to use $z_i = 1$. Thus, we consider the linear mixed model

$$y_i = X_i\beta + 1_{T_i}\gamma_i + \varepsilon_i, \tag{3.6}$$

where the random effects γ_i follow the same assumption as in (3.2), but unlike (3.2) the error components $\{\varepsilon_{it}\}$ for the given individual i are assumed to have an autocorrelation structure as in (2.23). That is,

$$\varepsilon_i \sim (0, A_i^{1/2} C_i A_i^{1/2}),$$

C_i being the $T_i \times T_i$ autocorrelation matrix as defined in (2.23). Furthermore, because $\text{var}(\varepsilon_{it}) = \sigma_\varepsilon^2$, for all $i = 1, \ldots, K$, and $t = 1, \ldots, T_i$, one may then write

$$\varepsilon_i \sim (0, \sigma_\varepsilon^2 C_i). \tag{3.7}$$

It then follows from (3.6) – (3.7) that

$$E[Y_{it}] = x_{it}'\beta$$

$$\text{var}[Y_{it}] = \sigma_\gamma^2 + \sigma_\varepsilon^2 = \sigma^2 \text{ (say)}$$

$$\text{cov}[Y_{iu}, Y_{it}] = \sigma_\gamma^2 + \sigma_\varepsilon^2 \rho_{|u-t|}, \tag{3.8}$$

yielding the mean and the covariance matrix of the response vector $y_i = (y_{i1}, \ldots, y_{iT_i})'$ as

$$E[Y_i] = X_i\beta, \quad \text{cov}[Y_i] = \Sigma_i = \sigma_\gamma^2 1_{T_i} 1_{T_i}' + \sigma_\varepsilon^2 C_i. \tag{3.9}$$

3.1.1 GLS Estimation of β

The β parameter is involved in the expectation of y_i in (3.9), therefore for known values of σ_γ^2, σ_ε^2, and ρ_ℓ ($\ell = 1, \ldots, T_i$), one may obtain the GLS estimate of β by using the formula

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{K} X_i' \Sigma_i^{-1} X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' \Sigma_i^{-1} y_i\right], \tag{3.10}$$

which is similar to the formula (2.13) for the GLS estimator of β under the linear fixed longitudinal model. The difference between (3.10) and (2.13) is that Σ_i in (2.13) has the form $\Sigma_i = \sigma_\varepsilon^2 C_i = \sigma^2 C_i$, whereas $\Sigma_i = \sigma_\gamma^2 1_{T_i} 1_{T_i}' + \sigma_\varepsilon^2 C_i$ in (3.10) with $\sigma_\gamma^2 + \sigma_\varepsilon^2 = \sigma^2$. Note that Σ_i^{-1} in (3.10) has the formula

$$\Sigma_i^{-1} = \frac{1}{\sigma_\varepsilon^2} C_i^{-1} - \frac{\sigma_\gamma^2}{\sigma_\varepsilon^4} \left[\frac{C_i^{-1} 1_{T_i} 1_{T_i}' C_i^{-1}}{1 + \frac{\sigma_\gamma^2}{\sigma_\varepsilon^4} 1_{T_i}' C_i^{-1} 1_{T_i}}\right], \tag{3.11}$$

which may be easily calculated once the inverse of the error correlation matrix C_i is known. Note that when the errors $\{\varepsilon_{it}\}$ in the mixed model (3.6) follow the general autocorrelation structure as in (2.23), one may easily obtain the C_i^{-1} matrix using any standard software such as *FORTRAN-90, R,* or *S-PLUS*. As discussed in Chapter 2, for specific AR(1) (2.32), MA(1) (2.35), and EQC (2.37) structures, C_i^{-1} may be calculated directly using the formulas given in Exercises 5, 6, and 7 of Chapter 2.

3.1.2 Moment Estimating Equations for σ_γ^2 and ρ_ℓ

For convenience we estimate

$$\phi = \frac{\sigma_\gamma^2}{\sigma_\gamma^2 + \sigma_\varepsilon^2} = \frac{\sigma_\gamma^2}{\sigma^2},\ \sigma^2,\ \text{and}\ \rho_\ell\ (\ell = 1,\ldots,T_i). \tag{3.12}$$

It is clear from (3.6) that

$$E[\sum_{i=1}^{K} \varepsilon_i' \varepsilon] = E[\sum_{i=1}^{K} (y_i - X_i\beta)'(y_i - X_i\beta)] = \sigma^2 \sum_{i=1}^{K} T_i,$$

where $\sigma^2 = \sigma_\gamma^2 + \sigma_\varepsilon^2$. Thus, we obtain a moment estimator for σ^2 as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{K} (y_i - X_i\hat{\beta}_{GLS})'(y_i - X_i\hat{\beta}_{GLS})}{\sum_{i=1}^{K} T_i}, \tag{3.13}$$

where $\hat{\beta}_{GLS}$ is given by (3.10). Note that the moment estimator $\hat{\sigma}^2$ in (3.13) is a consistent estimator for σ^2 as it is obtained from an unbiased moment estimating equation.

Next, we develop a moment estimating equation for $\phi = \sigma_\gamma^2/\sigma^2$ as follows. Similar to (2.40), suppose that δ_{it} is an indicator variable such that

$$\delta_{it} = \begin{cases} 1 & \text{if } t \leq T_i \\ 0 & \text{if } T_i < t \leq T. \end{cases}$$

for all $t = 1,\ldots,T$. Also, suppose that $d_i = (y_i - X_i\hat{\beta}_{GLS})$ and d_{it} denote the element of d_i corresponding to the tth $(t = 1,\ldots,T_i)$ element of the ith $(i = 1,\ldots,K)$ individual/cluster. By pooling the sample sum of squares and sum of products and equating to its population counterpart we obtain

$$\sum_{i=1}^{K} \sum_{u,t=1}^{T} \delta_{iu}\delta_{it} d_{iu} d_{it}/\sigma^2 = \phi \sum_{i=1}^{K} \sum_{u,t=1}^{T} \delta_{iu}\delta_{it}$$

$$+\ (1-\phi) \sum_{i=1}^{K} [T_i + 2\{(T_i - 1)\rho_1 + \ldots + 2\rho_{T_i-2} + \rho_{T_i-1}\}], \tag{3.14}$$

where $\rho_{|t-u|}$ is the $|t-u|$th lag autocorrelation used to define the general autocorrelation matrix C_i in (2.23). To solve (3.14) for ϕ and ρ_ℓ (ℓth lag autocorrelation), one may first write $\hat{\phi}$ as a function of ρ_ℓ as

$$\hat{\phi} = \frac{s - \sum_{i=1}^{K} [T_i + 2\{(T_i - 1)\rho_1 + \ldots + 2\rho_{T_i-2} + \rho_{T_i-1}\}]}{\sum_{i=1}^{K} \sum_{u,t=1}^{T} \delta_{iu}\delta_{it} - \sum_{i=1}^{K} [T_i + 2\{(T_i - 1)\rho_1 + \ldots + 2\rho_{T_i-2} + \rho_{T_i-1}\}]}, \tag{3.15}$$

where we write

$$s = \left[\sum_{i=1}^{K}\sum_{u,t=1}^{T} \delta_{iu}\delta_{it}d_{iu}d_{it}\right]\left[\sum_{i=1}^{K}\sum_{t=1}^{T}\delta_{it}d_{it}^2/\sum_{i=1}^{K}T_i\right]^{-1}. \tag{3.16}$$

As the ρ_ℓ values are involved in the covariance matrix Σ_i defined in (3.9), for an initial value of $\hat{\phi}$, say $\hat{\phi}_0$, we compute $\hat{\rho}_\ell$ as

$$\hat{\rho}_\ell = \frac{1}{1-\hat{\phi}_0}\left[\frac{\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\delta_{it}\delta_{i,t+\ell}d_{it}d_{i,t+\ell}/\{\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\delta_{it}\delta_{i,t+\ell}\}}{\sum_{i=1}^{K}\sum_{t=1}^{T}\delta_{it}d_{it}^2/\sum_{i=1}^{K}T_i} - \hat{\phi}_0\right] \tag{3.17}$$

[cf. Sneddon and Sutradhar (2004, eqn. (16)) in a more general linear familial longitudinal setup] for $\ell = 1,\ldots,T_i - 1$. Note that the initial value $\hat{\phi}_0$ in (3.17) may be computed by pretending $\rho_\ell = 0$ and then exploiting the off-diagonal elements of Σ_i. Thus, the formula for $\hat{\phi}_0$ is given by

$$\hat{\phi}_0 = \frac{\sum_{i=1}^{K}\sum_{u\neq t}^{T}\delta_{iu}\delta_{it}d_{iu}d_{it}/\sum_{i=1}^{K}\sum_{u\neq t}^{T}\delta_{iu}\delta_{it}}{\sum_{i=1}^{K}\sum_{t=1}^{T}\delta_{it}d_{it}^2/\sum_{i=1}^{K}T_i}. \tag{3.18}$$

Note the estimates of ϕ from (3.15) and of ρ_ℓ from (3.17) are then used to obtain improved estimates of β and σ^2 by (3.10) and (3.13), respectively. Next the improved estimates of β and σ^2 are used to obtain improved estimates of ϕ and ρ_ℓ. This constitutes a cycle of iterations which continues until convergence.

3.1.3 Linear Mixed Models for Rat Data

We reanalyze the rat data by using the linear longitudinal mixed model (3.6), whereas a longitudinal fixed model was used in Section 2.3 to analyze this rat dataset. In addition to the assumptions used for the fixed model, it has been assumed under the present mixed model that all 32 rats may have their own individual random effects γ_i $(i = 1,\ldots,32)$ with mean zero and variance σ_γ^2. Thus, if σ_γ^2 is found to be zero, then the mixed model would reduce to the fixed model. We now compute this variance of the random effects (σ_γ^2) along with the regression effects β in (3.6). We also compute the error variance σ_ε^2 and the lag correlations ρ_ℓ $(\ell = 1,\ldots,T-1)$ of the components of the error vector ε_i. Here $T = 6$. These estimates, by (3.8), provide the mean, variance, and correlations of the rat food data; that is,

$$E[Y_{it}] = x_{it}'\beta, \quad \text{var}(Y_{it}) = \sigma_\gamma^2 + \sigma_\varepsilon^2, \text{ and } \text{corr}(Y_{it}, Y_{i,t+\ell}) = \frac{\sigma_\gamma^2 + \sigma_\varepsilon^2\rho_\ell}{\sigma_\gamma^2 + \sigma_\varepsilon^2},$$

respectively. For convenience we estimate β by (3.10), $\sigma^2 = \sigma_\gamma^2 + \sigma_\varepsilon^2$ by (3.13), $\phi = \sigma_\gamma^2/\sigma^2$ by (3.15), and ρ_ℓ by (3.17), so that the estimates of σ_γ^2 and σ_ε^2 would be computed as

$$\hat{\sigma}_\gamma^2 = \hat{\phi}\hat{\sigma}^2 \text{ and } \hat{\sigma}_\varepsilon^2 = (1-\hat{\phi})\hat{\sigma}^2.$$

Applying Mixed Model to the Control Group Data

For the control group, the regression effects were found to be

$$\hat{\beta}_{c,0} = 9.05, \quad \hat{\beta}_{c,1} = 0.42,$$

with respective standard errors

$$\text{s.e.}(\hat{\beta}_{c,0}) = 0.45, \ \text{s.e.}(\hat{\beta}_{c,1}) = 0.45.$$

The estimates of ϕ and σ^2 were found to be $\hat{\phi} = 0.3275$ and $\hat{\sigma}^2 = 14.679$, leading to the estimates of σ_γ^2 and σ_ε^2 as

$$\hat{\sigma}_\gamma^2 = 4.808 \text{ and } \hat{\sigma}_\varepsilon^2 = 9.87,$$

respectively. Note that under the fixed model analysis, the variances for the data at different time points $(t = 1,\ldots,6)$ were found to range from 12.01 to 18.84. The estimate of the common variance under the mixed model, that is, $\hat{\sigma}^2 = 14.679$ appears to agree quite well with the variances computed under the fixed model setup. This in turn shows that the individual random effects variance estimate $\hat{\sigma}_\gamma^2 = 4.808$ is quite reasonable, and its large value indicates that the individual latent effects of the 32 rats are quite different. Thus, it may be much more reasonable to fit the mixed effects model to this dataset as compared to the use of the results obtained under the fixed model. The lag correlations for the errors were estimated as

$$\hat{\rho}_1 = 0.32, \quad \hat{\rho}_2 = -0.06, \quad \hat{\rho}_3 = -0.17, \quad \hat{\rho}_4 = -0.20, \quad \hat{\rho}_5 = -0.46.$$

To understand the lag correlations for the food eaten by the rats, we use the formula

$$\text{corr}(Y_{it}, Y_{i,t+\ell}) = \rho_\ell(y) = \frac{\sigma_\gamma^2 + \sigma_\varepsilon^2 \rho_\ell(\varepsilon)}{\sigma_\gamma^2 + \sigma_\varepsilon^2},$$

and obtain them as

$$\hat{\rho}_1 = 0.54, \ \hat{\rho}_2 = 0.29, \ \hat{\rho}_3 = 0.21, \ \hat{\rho}_4 = 0.19, \ \hat{\rho}_5 = 0.02,$$

which are in extremely good agreement with those computed under the fixed model, namely,

$$\hat{\rho}_1 = 0.55, \quad \hat{\rho}_2 = 0.31, \quad \hat{\rho}_3 = 0.22, \quad \hat{\rho}_4 = 0.17, \quad \hat{\rho}_5 = -0.01$$

(see Section 2.3).

Applying Mixed Model to the Treatment Groups Data

We now apply the longitudinal mixed model (3.6) to the treatment based amount of food eaten by 32 rats, and find the regression effects as

$$\hat{\beta}_{Tr,0} = 7.6552, \quad \hat{\beta}_{Tr,1} = 0.6018, \quad \hat{\beta}_{Tr,2} = 1.5728,$$

$$\hat{\beta}_{Tr,3} = -0.5959, \quad \hat{\beta}_{Tr,4} = -2.5328,$$

with respective standard errors

$$\text{s.e.}(\hat{\beta}_{Tr,0}) = 0.7085, \ \text{s.e.}(\hat{\beta}_{Tr,1}) = 0.3579, \ \text{s.e.}(\hat{\beta}_{Tr,2}) = 1.0020,$$

$$\text{s.e.}(\hat{\beta}_{Tr,3}) = 1.0065, \quad \text{s.e.}(\hat{\beta}_{Tr,4}) = 1.0022.$$

Note that these values differ slightly from the corresponding regression estimates in Chapter 2 found under the fixed model.

Next, the estimates of ϕ and σ^2 for the treatment group data are found to be $\hat{\phi} = 0.2212$ and $\hat{\sigma}^2 = 11.432$, leading to the estimates of σ_γ^2 and σ_ε^2 as

$$\hat{\sigma}_\gamma^2 = 2.529 \ \text{and} \ \hat{\sigma}_\varepsilon^2 = 8.903,$$

respectively. Note that under the fixed model analysis for the treatment based data, the variances for the data at different time points $(t = 1, \ldots, 6)$ were found to range from 7.38 to 14.26. The estimate of the common variance under the mixed model for the treatment based data, that is, $\hat{\sigma}^2 = 11.43$ appears to agree quite well with the variances computed under the fixed model setup. Note that as the random effects variance estimate $\hat{\sigma}_\gamma^2 = 2.529$ is far away from zero (even though it is smaller than the control data based value), it indicates that the individual latent effects of the 32 rats are quite different.

The lag correlations for the errors for the treatment based data were estimated as

$$\hat{\rho}_1 = 0.22, \ \hat{\rho}_2 = -0.08, \ \hat{\rho}_3 = 0.04, \ \hat{\rho}_4 = -0.23, \ \hat{\rho}_5 = -0.46.$$

By using the formula

$$\text{corr}(Y_{it}, Y_{i,t+\ell}) = \rho_\ell(y) = \frac{\sigma_\gamma^2 + \sigma_\varepsilon^2 \rho_\ell(\varepsilon)}{\sigma_\gamma^2 + \sigma_\varepsilon^2},$$

the lag correlations for the responses are found to be

$$\hat{\rho}_1 = 0.40, \quad \hat{\rho}_2 = 0.16, \quad \hat{\rho}_3 = 0.26, \quad \hat{\rho}_4 = 0.20, \quad \hat{\rho}_5 = -0.14,$$

respectively, which are in good agreement with the lag correlations found under the fixed model applied to the treatment based data.

3.2 Linear Dynamic Mixed Models for Balanced Longitudinal Data

In the econometrics literature [see, e.g., Amemiya (1985, Section 6.6.3); Hsiao (2003)], many authors have modelled the longitudinal dependence through an AR(1) type dynamic relationship between two lag 1 responses. Balestra and Nerlove (1966) used this type of dynamic model to analyze the demand for natural gas in 36 U.S.A. states in the period from 1950 to 1962. Bun and Carree (2005) also have used this type of first-order dynamic panel data to analyze unemployment rate data for ten years collected from 51 U.S.A. states. For simplicity, similar to these authors, we consider a balanced dynamic mixed model with $T_i = T$ for all $i = 1, \ldots, K$. This model is given by

$$y_{it} = x'_{it}\beta + \theta y_{i,t-1} + \gamma_i + \varepsilon_{it}, \tag{3.19}$$

where γ_i and ε_{it} satisfy the same assumptions

$$\gamma_i \overset{\text{iid}}{\sim} (0, \sigma_\gamma^2) \text{ and } \varepsilon_{it} \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2)$$

as for the mixed model (3.1). Thus, unlike the model (3.6), $\varepsilon_i = [\varepsilon_{i1}, \ldots, \varepsilon_{it}, \ldots, \varepsilon_{iT}]'$ in (3.19) now satisfies

$$\varepsilon_i \sim (01_T, \sigma_\varepsilon^2 I).$$

Note that in (3.19), θ represents the lag 1 dynamic dependence causing longitudinal correlations among the repeated responses. If the value of the initial response y_{i1} is known, then the mean of the response at time t depends on the covariate history as well as y_{i1}. To be specific, the mean under the model (3.19) has the form

$$E[Y_{it}] = \sum_{j=0}^{t-2} \theta^j x'_{i,t-j}\beta + \theta^{t-1} y_{i1}, \tag{3.20}$$

whereas the mean at time point t under the model (3.1) or (3.6) has the formula $E[Y_{it}] = x'_{it}\beta$, which depends on the covariate information for the time point t only. Recently, Rao, Sutradhar and Pandit (2009) have considered the dynamic dependence model given by

$$\begin{aligned} y_{i1} &= x'_{i1}\beta + \gamma_i + \varepsilon_{i1} \\ y_{it} &= x'_{it}\beta + \theta(y_{i,t-1} - x'_{i,t-1}\beta) + \gamma_i + \varepsilon_{it}, \text{ for } t = 2, \ldots, T, \end{aligned} \tag{3.21}$$

which produces the same mean, $E[Y_{it}] = x'_{it}\beta$, as that of the model (3.6). Note that in (3.21), the initial observation y_{i1} is assumed to be available through a random process similar to the rest of the observations. This assumption is more practical than assuming y_{i1} as fixed and given. See Hsiao (2003, Section 4.3.2, p. 76), for example, for this and other assumptions on the availability of the initial observation y_{i1}. We now provide below the basic mean, variance, and correlation properties of the model (3.21).

3.2.1 Basic Properties of the Dynamic Dependence Mixed Model (3.21)

We provide the first– and second-order moments based basic properties of the model (3.21) as in the following theorem.

Theorem 3.1. Under the dynamic mixed model (3.21), the mean and the variance of y_{it} $(t = 1, \ldots, T)$ are given by

$$E[Y_{it}] = \mu_{it} = x'_{it}\beta, \text{ and} \tag{3.22}$$

$$\text{var}[Y_{it}] = \sigma_{itt} = \sigma_\gamma^2 \left\{ \sum_{j=0}^{t-1} \theta^j \right\}^2 + \sigma_\varepsilon^2 \sum_{j=0}^{t-1} \theta^{2j}, \tag{3.23}$$

respectively, and the autocovariance at lag $t-u$ for $u < t$, is given by

$$\text{cov}[Y_{iu}, Y_{it}] = \sigma_\gamma^2 \sum_{j=0}^{t-1} \theta^j \sum_{k=0}^{u-1} \theta^k + \sigma_\varepsilon^2 \sum_{j=0}^{u-1} \theta^{t-u+2j}. \tag{3.24}$$

Proof: For all $t = 1, \ldots, T$, we first write

$$y_{it} - x'_{it}\beta = \sum_{j=0}^{t-1} \theta^j (\sigma_\gamma \gamma_i^* + \varepsilon_{i,t-j}), \tag{3.25}$$

where $\gamma_i^* = \gamma_i / \sigma_\gamma$. so that $\gamma_i^* \overset{iid}{\sim} (0,1)$. It then follows that

$$E(Y_{it} - x'_{it}\beta) = E_{\gamma_i^*} E[(Y_{it} - x'_{it}\beta) | \gamma_i^*] = 0, \tag{3.26}$$

and

$$E(y_{it} - x'_{it}\beta)^2 = E_{\gamma_i^*} E\left[\left\{ \sigma_\gamma \gamma_i^* \sum_{j=0}^{t-1} \theta^j + \Sigma \theta^j \varepsilon_{i,t-j} \right\}^2 | \gamma_i^* \right]$$

$$= \sigma_\gamma^2 \left[\sum_{j=0}^{t-1} \theta^j \right]^2 + \sigma_\varepsilon^2 \sum_{j=0}^{t-1} \theta^{2j}, \tag{3.27}$$

yielding the mean and the variance of y_{it} as in the theorem. Next for $u < t$, it follows from (3.25) that the covariance between y_{iu} and y_{it} is given by

$$\sigma_{iut} = \text{cov}(Y_{iu}, Y_{it})$$

$$= E(Y_{iu} - x'_{iu}\beta)(Y_{it} - x'_{it}\beta)$$

$$= E_{\gamma_i^*}[E\{(Y_{iu} - x'_{iu}\beta)(Y_{it} - x'_{it}\beta)\}|\gamma_i^*]$$

$$= E_{\gamma_i^*}\left[E\left\{\sum_{j=0}^{u-1}\theta^j(\sigma_\gamma\gamma_i^* + \varepsilon_{i,u-j})\sum_{j=0}^{t-1}\theta^j(\sigma_\gamma\gamma_i^* + \varepsilon_{i,t-j})\right\}|\gamma_i^*\right]$$

$$= \sigma_\gamma^2\sum_{j=0}^{t-1}\theta^j\sum_{k=0}^{u-1}\theta^k + \sigma_\varepsilon^2\sum_{j=0}^{u-1}\theta^{t-u+2j}, \quad (3.28)$$

which is the same as equation (3.24).

Note that for the estimation of θ by the method of moments, it is sufficient to exploit all lag 1 pairwise products. This prompts us to write the lag 1 autocovariance under the model (3.21) as in the following corollary.

Corollary 3.1. For $t - u = 1$, the lag 1 autocovariance is given by

$$\begin{aligned}\text{cov}[Y_{it}, Y_{i,t+1}] = \sigma_{it,t+1} &= \theta\left[\sigma_\gamma^2\{\sum_{j=0}^{t-1}\theta^j\}^2 + \sigma_\varepsilon^2\sum_{j=0}^{t-1}\theta^{2j}\right]\\ &= \theta\sigma_{itt}. \quad (3.29)\end{aligned}$$

3.2.2 Estimation of the Parameters of the Dynamic Mixed Model (3.21)

a. Least Square Dummy Variable (LSDV) Estimator

LSDV Estimation of θ and β

Rewrite the model (3.21) as

$$\begin{aligned}y_{i1} &= x'_{i1}\beta + \gamma_i + \varepsilon_{i1},\\ y_{it} &= \theta y_{i,t-1} + (x_{it} - x_{i,t-1})'\beta + \gamma_i + \varepsilon_{it}\\ &= \theta y_{i,t-1} + w'_{it}\beta + \gamma_i + \varepsilon_{it}, \text{ for } t = 2,\ldots,T. \quad (3.30)\end{aligned}$$

Define

$$\bar{y}_i = \frac{1}{T-1}\sum_{t=2}^{T} y_{it},\ \bar{y}_{i,-1} = \frac{1}{T-1}\sum_{t=2}^{T} y_{i,t-1},\ \bar{x}_i = \frac{1}{T-1}\sum_{t=2}^{T} x_{it},$$

$$\bar{x}_{i,-1} = \frac{1}{T-1}\sum_{t=2}^{T} x_{i,t-1},\quad \bar{w}_i = \bar{x}_i - \bar{x}_{i-1},\ \bar{\varepsilon}_i = \frac{1}{T-1}\sum_{t=2}^{T}\varepsilon_{it},$$

and

$$\tilde{y}_{it} = y_{it} - \bar{y}_i, \quad \tilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}, \quad \tilde{w}_{it} = w_{it} - \bar{w}_i, \;\; \tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i,$$

and rewrite the model (3.30) as

$$\tilde{y}_{it} = \theta \tilde{y}_{i,t-1} + \tilde{w}'_{it}\beta + \tilde{\varepsilon}_{it}, \;\; \text{for } t = 2, \ldots, T, \tag{3.31}$$

which is free from γ_i. The LSDV estimators of θ and β are obtained by applying the OLS (ordinary least squares) method to (3.31) [Bun and Carree (2005, Section 2); see also Hsiao (2003, Section 4.2)]. Let $\hat{\theta}_{lsdv}$ and $\hat{\beta}_{lsdv}$ denote the LSDV estimators of θ and β, respectively. By writing $x^*_{it} = (\tilde{y}_{i,t-1}, \tilde{w}'_{it})'$, and using the notation

$$\tilde{y}_i = [\tilde{y}_{i2}, \ldots, \tilde{y}_{iT}]' : (T-1) \times 1, \;\; \text{and} \;\; X^*_i = [x^*_{i2}, \ldots, x^*_{it}, \ldots, x^*_{iT}]' : (T-1) \times (p+1),$$

where p is the dimension of the β vector, the LSDV estimators have the formula given by

$$\begin{bmatrix} \hat{\theta}_{lsdv} \\ \\ \hat{\beta}_{lsdv} \end{bmatrix} = \left[\sum_{i=1}^{K} X^{*\prime}_i X^*_i \right]^{-1} \left[\sum_{i=1}^{K} X^{*\prime}_i \tilde{y}_i \right]. \tag{3.32}$$

These LSDV estimators are known to be biased and hence inconsistent for the respective true parameters. Bun and Carree (2005) have discussed a bias correction approach for a dynamic mixed model with scalar β ($p = 1$) and provided the bias-corrected LSDV (BCLSDV) estimator of θ and β as

$$\hat{\theta}_{bclsdv} = \hat{\theta}_{lsdv} + \frac{\hat{\sigma}^2_{\varepsilon} h(\hat{\theta}_{bclsdv}, T-1)}{(1 - \hat{\rho}^2_{wy_{-1}})\hat{\sigma}^2_{y_{-1}}}, \quad \hat{\beta}_{bclsdv} = \hat{\beta}_{lsdv} + \hat{\xi}(\hat{\theta}_{lsdv} - \hat{\theta}_{bclsdv}), \tag{3.33}$$

[Bun and Carree (2005, eqn. (13), p. 13)] where

$$h(\theta, T-1) = \frac{(T-2) - (T-1)\theta + \theta^{T-1}}{(T-1)(T-2)(1-\theta)^2}$$

$$\hat{\rho}_{wy_{-1}} = \frac{\hat{\sigma}_{wy_{-1}}}{\hat{\sigma}_w \hat{\sigma}_{y_{-1}}}$$

$$\hat{\xi} = \frac{\hat{\sigma}_{wy_{-1}}}{\hat{\sigma}^2_w}, \tag{3.34}$$

with

$$\hat{\sigma}^2_{y_{-1}} = \frac{1}{K(T-2)} \sum_{i=1}^{K} \sum_{t=2}^{T} \tilde{y}^2_{i,t-1}, \quad \hat{\sigma}^2_w = \frac{1}{K(T-2)} \sum_{i=1}^{K} \sum_{t=2}^{T} \tilde{w}^2_{it}$$

$$\hat{\sigma}_{wy_{-1}} = \frac{1}{K(T-2)} \sum_{i=1}^{K} \sum_{t=2}^{T} \tilde{w}_{it} \tilde{y}_{i,t-1}, \tag{3.35}$$

where K is assumed to be large. Note that for the bias-correction estimation, it is required to have $T \geq 3$.

b. Instrumental Variable (IV) Estimator

IV Estimation of θ and β

Note that the model (3.31) derived from (3.30) is free of γ_i. To avoid the estimation of γ_i or say σ_γ^2 in (3.30), many econometricians have considered an alternative dynamic model utilizing the first difference of the responses [e.g. Hsiao (2003, Section 4.3.3.c)] as

$$y_{it} - y_{i,t-1} = \theta(y_{i,t-1} - y_{i,t-2}) + (w_{it} - w_{i,t-1})'\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}), \;\; \text{for } t = 3, \ldots, T, \tag{3.36}$$

y_{i1} being the initial response. Now any variable such as

$$[y_{i,t-2-j} - y_{i,t-3-j}] \;\; \text{for } j = 0, 1, \ldots$$

is referred to as an instrumental variable for $[y_{i,t-1} - y_{i,t-2}]$ provided

$$E[(Y_{i,t-1} - Y_{i,t-2})(Y_{i,t-2-j} - Y_{i,t-3-j})] \neq 0,$$

but

$$E[(\varepsilon_{it} - \varepsilon_{i,t-1})(Y_{i,t-2-j} - Y_{i,t-3-j})] = 0.$$

Suppose that for simplicity we consider only one instrumental variable $y_{i,t-2-j} - y_{i,t-3-j}$ with $j = 0$ and estimate θ and β for the model (3.36). Write $x_{it}^* = ((y_{i,t-1} - y_{i,t-2}), (w_{it} - w_{i,t-1})')'$ and define

$$X_i^* = [x_{i4}^*, \ldots, x_{it}^*, \ldots, x_{iT}^*]' : (T-3) \times (p+1),$$

where p is the dimension of the β vector. Now by using the instrumental variable, write $s_{it}^* = ((y_{i,t-2} - y_{i,t-3}), (w_{it} - w_{i,t-1})')'$ and define

$$S_i^* = [s_{i4}^*, \ldots, s_{it}^*, \ldots, s_{iT}^*]' : (T-3) \times (p+1).$$

Further define

$$y_i^* = [y_{i4} - y_{i3}, \ldots, y_{iT} - y_{i,T-1}]' : (T-3) \times 1.$$

One then obtains the IV estimates of θ and β by using the formula

$$\begin{bmatrix} \hat{\theta}_{iv} \\ \\ \hat{\beta}_{iv} \end{bmatrix} = \left[\sum_{i=1}^{K} S^{*\prime}_i X_i^* \right]^{-1} \left[\sum_{i=1}^{K} S^{*\prime}_i y_i^* \right], \tag{3.37}$$

[Amemiya (1985, p. 11 – 12)]. Note that in this approach it is required to have $T \geq 4$, which can be a major limitation. This is because, in practice, in the longitudinal/panel data setup, T may be as small as 2.

c. IV Based Generalized Method of Moments (GMM) Estimators

IV Based GMM Estimation of θ and β

Note that $y_{i,t-2-j}$ for $j = 0, 1, \ldots, t-3$ are also instrumental variables for $[y_{i,t-1} - y_{i,t-2}]$ as

$$E[(Y_{i,t-1} - Y_{i,t-2})Y_{i,t-2-j}] \neq 0, \text{ but } E[(\varepsilon_{it} - \varepsilon_{i,t-1})Y_{i,t-2-j}] = 0.$$

Define

$$q_{it} = [y_{i1}, \ldots, y_{i,t-2}, w_i']', \text{ where } w_i' = [w_{i2}', \ldots, w_{iT}']'.$$

The following moment conditions are satisfied:

$$E[q_{it}u_{it}] = 0, \text{ for } t = 3, \ldots, T, \tag{3.38}$$

where $u_{it} = \varepsilon_{it} - \varepsilon_{i,t-1} = (y_{it} - y_{i,t-1}) - \theta(y_{i,t-1} - y_{i,t-2}) - (w_{it} - w_{i,t-1})'\beta$. Let $u_i = [u_{i3}, \ldots, u_{iT}]'$ be the $(T-2) \times 1$ vector of the first difference of errors. All possible moment conditions in (3.38) may be then represented by

$$E[Q_i u_i] = 0, \tag{3.39}$$

where Q_i is the $s \times (T-2)$ diagonal matrix given by

$$Q_i = \begin{bmatrix} q_{i3} & 0 & 0 \cdots & 0 \\ 0 & q_{i4} & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \cdots & q_{iT} \end{bmatrix}. \tag{3.40}$$

The GMM estimator of $\alpha = (\theta, \beta')'$ proposed by Arellano and Bond (1991) [see also Hansen (1982)] is obtained by minimizing

$$\left[\frac{1}{K}\sum_{i=1}^{K} u_i'Q_i'\right] \Psi^{-1} \left[\frac{1}{K}\sum_{i=1}^{K} Q_i u_i\right], \tag{3.41}$$

where

$$\Psi = E\left[\frac{1}{K^2}\sum_{i=1}^{K} Q_i u_i u_i' Q_i'\right].$$

Thus, the IV based GMM estimating equation for α is given by

$$\left[\frac{1}{K}\sum_{i=1}^{K}\frac{\partial u_i'}{\partial \alpha}Q_i'\right]\Psi^{-1}\left[\frac{1}{K}\sum_{i=1}^{K}Q_iu_i\right]=0. \tag{3.42}$$

d. Some Remarks on Moment Estimation of σ_ε^2 and σ_γ^2

Note that all three estimation methods, namely LSDV, IV, and IV based GMM methods are developed in such a way that one can estimate the regression effects β and the dynamic dependence parameter θ without estimating σ_ε^2 and γ_i that lead to the estimate of σ_γ^2. In many situations in practice, the estimation of the variance components σ_γ^2 and σ_ε^2 may also be of interest. For example, to develop the bias-corrected LSDV estimators of β and θ, one needs the estimate of σ_ε^2 [see Exercise 3.3; see also Bun and Carree (2005)]. As far as the variance parameter σ_γ^2 is concerned, it is sometimes of direct interest to explain the variation that may be present among the individuals or individual firms. However, it may not be easy to estimate these parameters, especially σ_γ^2, consistently. Some authors have used the well-known ordinary method of moments [see Hsiao (2003, eqns. (4.3.35) and (4.3.36)), e.g.) to achieve this goal, but problems arise when T is small (e.g., $T=2,3$) which is in fact a more realistic case in the panel and/or longitudinal data setup. Because the LSDV, IV, and IV based GMM approaches are developed based on the first difference variables (or variables deviated from the mean of the individual group), their unbiasedness and consistency for the estimation of β and θ are also affected in cases when T is small.

For the sake of completeness, we provide the so-called moment estimators for the σ_ε^2 and σ_γ^2 [Hsiao (2003, eqns. (4.3.35) and (4.3.36))] as

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{K}\sum_{t=3}^{T}[(y_{it}-y_{i,t-1})-\hat{\theta}(y_{i,t-1}-y_{i,t-2})-\hat{\beta}'(w_{it}-w_{i,t-1})]^2}{2K(T-2)}, \tag{3.43}$$

$$\hat{\sigma}_\gamma^2 = \frac{\sum_{i=1}^{K}[\bar{y}_i-\hat{\theta}\bar{y}_{i,-1}-\hat{\beta}'\bar{w}_i]^2}{K}-\frac{1}{T-1}\hat{\sigma}_\varepsilon^2, \tag{3.44}$$

where $\bar{y}_i$, $\bar{y}_{i,-1}$, and $\bar{w}_i$ are given in (3.31), and $\hat{\beta}$ and $\hat{\theta}$ represent any of the LSDV, IV, or IV based GMM estimates.

3.3 Further Estimation for the Parameters of the Dynamic Mixed Model

In this section, we provide two new estimation procedures, following Rao, Sutradhar, and Pandit (2010). The first procedure is an improvement over the well-known MM (method of moments) and may be referred to as the improved MM (IMM) approach. See, for example, Sutradhar (2004) and Jiang and Zhang (2001), for such an IMM approach in the context of nonlinear, namely binary and count data analysis. Alternatively, similar to Rao, Sutradhar, and Pandit (2010), this IMM approach may also be referred to as the GMM approach, which, however, unlike the IV based GMM approach discussed in the econometrics literature (see previous section), uses the IV concept indirectly. As far as the properties of the IMM/GMM and MM approaches are concerned, both IMM/GMM and MM approaches produce consistent estimates for the parameters of the model, but MM estimates are less efficient than the IMM/GMM estimates. The new GMM approach (as compared to the IV based GMM approach) is given in the following subsection. In Section 3.3.2, we provide the second procedure, namely a generalized quasi-likelihood (GQL) approach that produces even more efficient estimates than the GMM/IMM approach.

Note that we discuss both GMM/IMM and GQL estimation procedures for a wider model than (3.21). Suppose that an additional fixed covariate z_i corresponding to the random effect γ_i is available from the ith $(i = 1,\ldots,K)$ individual. We then rewrite the model (3.21) as

$$\begin{aligned} y_{i1} &= x'_{i1}\beta + z_i\gamma_i + \varepsilon_{i1} \\ y_{it} &= x'_{it}\beta + \theta(y_{i,t-1} - x'_{i,t-1}\beta) + z_i\gamma_i + \varepsilon_{it}, \text{ for } t = 2,\ldots,T, \end{aligned} \tag{3.45}$$

just by inserting z_i as the coefficient of γ_i. The definition and the assumptions for other variables and parameters remain the same as in (3.21). Thus, if $z_i = 1$ for all $i = 1,\ldots,K$, then the linear dynamic mixed model (3.45) reduces to the model (3.21).

3.3.1 GMM/IMM Estimation Approach

Theorem 3.2. Under the dynamic mixed model (3.45), the mean and variance of y_{it} $(t = 1,\ldots,T)$ are given by

$$E[Y_{it}] = \mu_{it} = x'_{it}\beta, \text{ and} \tag{3.46}$$

$$\text{var}[Y_{it}] = \sigma_{itt} = z_i^2\sigma_\gamma^2 \left\{\sum_{j=0}^{t-1}\theta^j\right\}^2 + \sigma_\varepsilon^2\sum_{j=0}^{t-1}\theta^{2j}, \tag{3.47}$$

respectively, and the autocovariance at lag $t-u$ for $u < t$, is given by

$$\text{cov}[Y_{iu}, Y_{it}] = \sigma_{iut} = z_i^2\sigma_\gamma^2\sum_{j=0}^{t-1}\theta^j\sum_{k=0}^{u-1}\theta^k + \sigma_\varepsilon^2\sum_{j=0}^{u-1}\theta^{t-u+2j}. \tag{3.48}$$

Proof: The proof is similar to that of Theorem 3.1.

It is of interest to estimate all parameters of the model (3.45), namely,

$$\beta,\ \theta,\ \sigma_\gamma^2, \text{ and } \sigma_\varepsilon^2.$$

In estimating these parameters by the MM, we may construct four suitable distance functions by taking the difference between appropriate sample quantities and their corresponding population counterparts from (3.46) – (3.48). We write these distance functions as

$$\text{For } \beta : \psi_1 = \sum_{i=1}^{I} \sum_{t=1}^{T} x_{it}[y_{it} - x'_{it}\beta] \tag{3.49}$$

$$\text{For } \theta : \psi_2 = \sum_{i=1}^{I} \sum_{t=1}^{T-1} [\{(y_{it} - x'_{it}\beta)(y_{i,t+1} - x'_{i,t+1}\beta)\} - \sigma_{it,t+1}] \tag{3.50}$$

$$\text{For } \sigma^2_\gamma : \psi_3 = \sum_{i=1}^{I} \sum_{u<t}^{T} [\{(y_{iu} - x'_{iu}\beta)(y_{it} - x'_{it}\beta)\} - \sigma_{iut}] \tag{3.51}$$

$$\text{For } \sigma^2_\varepsilon : \psi_4 = \sum_{i=1}^{I} \sum_{t=1}^{T} [\{y_{it} - x'_{it}\beta\}^2 - \sigma_{itt}]/IT$$
$$-2\sum_{i=1}^{I} \sum_{u<t}^{T} [\{(y_{iu} - x'_{iu}\beta)(y_{it} - x'_{it}\beta)\} - \sigma_{iut}]/IT(T-1). \tag{3.52}$$

Because $E(\psi) = E[\psi'_1,\ \psi_2,\ \psi_3,\ \psi_4]' = [01'_p, 0, 0, 0]'$, in the MM approach, one would have solved the four MM equations

$$\psi'_1 = 0, \quad \psi_2 = 0, \quad \psi_3 = 0, \quad \psi_4 = 0,$$

to obtain the MM estimates for $\beta,\ \theta,\ \sigma^2_\gamma$, and σ^2_ε.

Now by following the suggestion of Hansen (1982) [see also Jiang and Zhang (2001)], one may obtain the so-called GMM/IMM estimate for $\alpha = [\beta',\ \theta,\ \sigma^2_\gamma,\ \sigma^2_\varepsilon]'$ by minimizing the quadratic form

$$Q = \psi' V \psi \tag{3.53}$$

for a suitable $(p+3)\times(p+3)$, positive definite matrix V, with $V = [\text{cov}(\psi)]^{-1}$ as an optimal choice. Note, however, that because the computation of the cov(ψ) matrix requires the formulas for the third– and fourth-order moments as well, one cannot compute such a covariance matrix provided the error distributions for the model (3.45) are known. Furthermore, as the consistency of the estimators is not affected by the choice of the weight matrix, a possible resolution is to use a normality based matrix V_N, and solve the estimating equation

$$\frac{\partial \psi'}{\partial \alpha} V_N \psi = 0. \tag{3.54}$$

This estimating equation may be solved by using the Gauss–Newton iterative equation

$$\hat{\alpha}_{GMM}(r+1) = \hat{\alpha}_{GMM}(r) + \left[\frac{\partial \psi'}{\partial \alpha} V_N \frac{\partial \psi}{\partial \alpha'}\right]_r^{-1} \left[\frac{\partial \psi'}{\partial \alpha} V_N \psi\right]_r, \tag{3.55}$$

where $[\,]_r$ denotes that the expression within the square bracket is evaluated at $\alpha = \hat{\alpha}_{GMM}(r)$, the estimate obtained for the rth iteration. Let the final solution obtained from (3.55) be denoted by $\hat{\alpha}_{GMM}$. Under some mild regularity condition it may be shown that as $K \to \infty$,

$$K^{\frac{1}{2}}(\hat{\alpha}_{GMM} - \alpha) \sim N\left[0, K\left\{\frac{\partial \psi'}{\partial \alpha} V_N \frac{\partial \psi}{\partial \alpha}\right\}^{-1} \left(\frac{\partial \psi'}{\partial \alpha} V_N V^{-1} V_N \frac{\partial \psi}{\partial \alpha'}\right) \times \left\{\frac{\partial \psi'}{\partial \alpha} V_N \frac{\partial \psi}{\partial \alpha'}\right\}^{-1}\right], \tag{3.56}$$

where $V^{-1} = \text{cov}(\psi)$ is the true covariance matrix for ψ based on the true data such as Gaussian or elliptic or any other symmetric continuous data. Note that if the true distributions of the errors are normal, then $V = V_N$. This leads to the covariance matrix of $\hat{\alpha}_{GMM}$ as

$$\text{cov}(\hat{\alpha}_{GMM}) = \left\{\frac{\partial \psi'}{\partial \alpha} V_N \frac{\partial \psi}{\partial \alpha'}\right\}^{-1}. \tag{3.57}$$

Computation of $\partial \psi'/\partial \alpha$ in (3.54)

In order to compute the derivatives of the elements of ψ with respect to the elements of α for (3.54), we treat the β parameter in ψ_1 as unknown but it is known in ψ_2, ψ_3, and ψ_4. This is a reasonable treatment as the estimation is done by iteration. Now, the derivatives in (3.54) are easily obtained by using the formulas for the derivatives of $\mu_{it} = x'_{it}\beta$, σ_{itt}, and σ_{iut} with respect to the elements of $\alpha = [\beta', \theta, \sigma^2_\gamma, \sigma^2_\varepsilon]'$. These formulas are:

$$\frac{\partial \mu_{it}}{\partial \beta} = x_{it}, \frac{\partial \mu_{it}}{\partial \theta} = \frac{\partial \mu_{it}}{\partial \sigma^2_\gamma} = \frac{\partial \mu_{it}}{\partial \sigma^2_\varepsilon} = 0, \tag{3.58}$$

$$\frac{\partial \sigma_{itt}}{\partial \beta} = \frac{\partial \sigma_{iut}}{\partial \beta} = 0, \tag{3.59}$$

$$\frac{\partial \sigma_{itt}}{\partial \theta} = 2z_i^2 \sigma^2_\gamma \sum_{j=0}^{t-1} \theta^j \sum_{j=1}^{t-1} j\theta^{j-1} + 2\sigma^2_\varepsilon \sum_{j=1}^{t-1} j\theta^{2j-1}, \tag{3.60}$$

$$\frac{\partial \sigma_{iut}}{\partial \theta} = z_i^2 \sigma^2_\gamma \left[\sum_{j=0}^{t-1} \theta^j \sum_{k=1}^{u-1} k\theta^{k-1} + \sum_{j=1}^{t-1} j\theta^{j-1} \sum_{k=0}^{u-1} \theta^k\right]$$

$$+\sigma_\varepsilon^2 \sum_{j=0}^{u-1}(t-u+2j)\theta^{t-u+2j-1}, \tag{3.61}$$

$$\frac{\partial \sigma_{itt}}{\partial \sigma_\gamma^2} = z_i^2 \left\{\sum_{j=0}^{t-1}\theta^j\right\}^2, \quad \frac{\partial \sigma_{iut}}{\partial \sigma_\gamma^2} = z_i^2 \sum_{j=0}^{t-1}\theta^j \sum_{k=0}^{u-1}\theta^k, \tag{3.62}$$

and

$$\frac{\partial \sigma_{itt}}{\partial \sigma_\varepsilon^2} = \sum_{j=0}^{t-1}\theta^{2j}, \quad \frac{\partial \sigma_{iut}}{\partial \sigma_\varepsilon^2} = \sum_{j=0}^{u-1}\theta^{t-u+2j}. \tag{3.63}$$

Construction of V_N, the 'Working' Weight Matrix Under Normality

Note that $V = [\mathrm{cov}(\psi)]^{-1}$, where

$$\mathrm{cov}(\psi) = \begin{bmatrix} \mathrm{var}(\psi_1) & \mathrm{cov}(\psi_1,\psi_2) & \mathrm{cov}(\psi_1,\psi_3) & \mathrm{cov}(\psi_1,\psi_4) \\ & \mathrm{var}(\psi_2) & \mathrm{cov}(\psi_2,\psi_3) & \mathrm{cov}(\psi_2,\psi_4) \\ & & \mathrm{var}(\psi_3) & \mathrm{cov}(\psi_3,\psi_4) \\ & & & \mathrm{var}(\psi_4) \end{bmatrix}, \tag{3.64}$$

where ψ_1, ψ_2, ψ_3, and ψ_4 are given by (3.49), (3.50), (3.51), and (3.52), respectively. Further note that because the errors in the underlying model (3.45) do not have any specific distributions, it is in general not possible to compute the covariance matrix $\mathrm{cov}(\psi)$ as its computation requires the formulas for the third− and fourth-order moments of the data. The use of a suitable weight matrix may increase the efficiency of the estimator of α without affecting its consistency, thus Rao, Sutradhar, and Pandit (2010) used a 'working' normality based V matrix which we denote here by V_N. We remark that in constructing V_N, the true first− and second-order moments of the data will be used. Thus, this pretense of the normal distribution helps to compute the third− and fourth-order moments by using the true first− and second-order moments. Recall these first− and second-order moments from Theorem 3.2 and write the true covariance matrix of the data following the model (3.45) as

$$\Sigma_i = (\sigma_{iut}) = (\sigma_{itu}), \tag{3.65}$$

where the formulas for σ_{iut} are given in (3.47) – (3.48).

Now, under the 'working' normality pretention, the response vector

$$y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$$

follows a T-dimensional multinormal distribution with true mean vector $\mu_i = (\mu_{i1}, \ldots, \mu_{it}, \ldots, \mu_{iT})'$, and the $T \times T$ true covariance matrix $\Sigma_i = (\sigma_{iut})$, where $\mu_{it} = x_{it}'\beta$ as in (3.46) and Σ_i is given by (3.65) [see also (3.47) – (3.48)]. Note that because of this 'working' normality assumption, we may obtain the third− and the fourth-order moments of the responses using the following two lemmas.

Lemma 3.1. Under normality, the third-order moments are given as

$$\delta_{iu\ell t} = E\left[(Y_{iu}-\mu_{iu})(Y_{i\ell}-\mu_{i\ell})(Y_{it}-\mu_{it})\right] = 0. \tag{3.66}$$

Lemma 3.2. Under normality, the fourth-order moments have the formulas

$$\begin{aligned}\phi_{iu\ell mt} &= E\left[(Y_{iu}-\mu_{iu})(Y_{i\ell}-\mu_{i\ell})(Y_{im}-\mu_{im})(Y_{it}-\mu_{it})\right] \\ &= \sigma_{iu\ell}\sigma_{imt} + \sigma_{ium}\sigma_{i\ell t} + \sigma_{iut}\sigma_{i\ell m},\end{aligned} \tag{3.67}$$

where σ_{iut} are the true covariances given by (3.65).

By using (3.65) and (3.66), one obtains the variance of ψ_1, and its covariances with $\psi_2,\ \psi_3$, and ψ_4 as

$$\begin{aligned}\mathrm{var}_N(\psi_1) &= \sum_{i=1}^{I}\sum_{u=1}^{T}\sum_{t=1}^{T}\sigma_{iut}x_{iu}x_{it}' \\ \mathrm{cov}_N(\psi_1,\psi_2) &= \mathrm{cov}_N(\psi_1,\psi_3) = \mathrm{cov}_N(\psi_1,\psi_4) = 0.\end{aligned} \tag{3.68}$$

Next by using $(3.65)-(3.67)$, we obtain the variance of ψ_2 and its covariances with ψ_3 and ψ_4 as

$$\begin{aligned}\mathrm{var}_N(\psi_2) &= \sum_{i=1}^{I}\sum_{u=1}^{T-1}\sum_{t=1}^{T-1}\left[\phi_{iu(u+1)t(t+1)} - \sigma_{iu(u+1)}\sigma_{it(t+1)}\right] \\ \mathrm{cov}_N(\psi_2,\psi_3) &= \sum_{i=1}^{I}\sum_{u=1}^{T-1}\sum_{m<t}^{T}\left[\phi_{iu,u+1,mt} - \sigma_{iu,u+1}\sigma_{imt}\right] \\ \mathrm{cov}_N(\psi_2,\psi_4) &= (IT)^{-1}\sum_{i=1}^{I}\sum_{u=1}^{T-1}\sum_{t=1}^{T}\left[\phi_{iu(u+1)tt} - \sigma_{iu(u+1)}\sigma_{itt}\right] \\ &\quad -2\{IT(T-1)\}^{-1}\sum_{i=1}^{I}\sum_{u=1}^{T-1}\sum_{m<t}^{T}\left[\phi_{iu(u+1)mt} - \sigma_{iu(u+1)}\sigma_{imt}\right]\end{aligned} \tag{3.69}$$

Similarly, the variance of ψ_3 and the covariance between ψ_3 and ψ_4 are given by

$$\mathrm{var}_N(\psi_3) = \sum_{i=1}^{I}\sum_{u<\ell}^{T}\sum_{m<t}^{T}\left[\phi_{iu\ell mt} - \sigma_{iu\ell}\sigma_{imt}\right], \tag{3.70}$$

and

$$\mathrm{cov}_N(\psi_3,\psi_4) = (IT)^{-1}\sum_{i=1}^{I}\sum_{u=1}^{T}\sum_{m<t}^{T}[\phi_{iuumt}-\sigma_{iuu}\sigma_{imt}]$$

$$-2\{IT(T-1)\}^{-1}\sum_{i=1}^{I}\sum_{u<\ell}^{T}\sum_{m<t}^{T}[\phi_{iu\ell mt}-\sigma_{iu\ell}\sigma_{imt}], \quad (3.71)$$

respectively, and the variance of ψ_4 has the formula

$$\mathrm{var}_N(\psi_4) = (IT)^{-2}\sum_{i=1}^{I}\sum_{u=1}^{T}\sum_{t=1}^{T}[\phi_{iuutt}-\sigma_{iuu}\sigma_{itt}]$$

$$-2(IT)^{-1}(IT(T-1))^{-1}\sum_{i=1}^{I}\sum_{u=1}^{T}\sum_{m<t}^{T}[\phi_{iuumt}-\sigma_{iuu}\sigma_{imt}]$$

$$+4(IT(T-1))^{-2}\sum_{i=1}^{I}\sum_{u<\ell}^{T}\sum_{m<t}^{T}[\phi_{iu\ell mt}-\sigma_{iu\ell}\sigma_{imt}]. \quad (3.72)$$

This completes the construction of the V_N matrix.

3.3.2 GQL Estimation Approach

Note that in an independence setup when the responses follow an exponential family of distributions such as normal, binary and Poisson, Wedderburn (1974) [see also McCullagh (1983)] proposed a quasi-likelihood approach for independent data which exploits both the mean and variance in estimating the parameters such as β involved in the means of the responses. Thus, if the responses were following an independent model, say with $\theta = 0$, and all $\gamma_i = 0$ in (3.45), then the QL estimate for β involved in the means would be the solution of

$$\sum_{i=1}^{K}\frac{\partial \mu_i'}{\partial\beta}[\mathrm{diag}\{\mathrm{var}(\varepsilon_{i1}),\ldots,\mathrm{var}(\varepsilon_{it}),\ldots,\mathrm{var}(\varepsilon_{iT})\}]^{-1}(y_i-\mu_i)=0, \quad (3.73)$$

where $y_i = (y_{i1},\ldots,y_{iT})'$ is the $T\times 1$ vector of first order responses for the ith individual, and $\mu_i = E(Y_i) = [\mu_{i1},\ldots,\mu_{it},\ldots,\mu_{iT}]'$ is the mean vector with $\mu_{it} = x_{it}'\beta$, as given by the model (3.45). Because $\mathrm{var}(\varepsilon_{it}) = \sigma_\varepsilon^2$ for all $t = 1,\ldots,T$, the QL equation (3.73) reduces to

$$\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{K}\frac{\partial\mu_i'}{\partial\beta}I_T(y_i-\mu_i)=0,$$

which is in fact the ordinary least squares estimating equation yielding the independence based QL (QL(I)) or OLS estimator of β given by

$$\hat{\beta}_{QL(I)} = \hat{\beta}_{OLS} = [\sum_{i=1}^{K} X_i' X_i]^{-1} \sum_{i=1}^{K} X_i' y_i,$$

[see also eqn. (2.10)] where X_i is the $p \times T$ covariate matrix.

a. GQL/GLS Estimating Equation for β

Sutradhar (2003) generalized the QL estimation for the independence data to the correlated data setup for complex discrete data such as binary and count data. This still can be applied to the linear mixed model (3.45) and the generalized quasi-likelihood (GQL) estimating equation is given by

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1} (y_i - \mu_i) = 0, \qquad (3.74)$$

where Σ_i is the covariance matrix of y_i which contains all scale parameters, namely θ, σ_γ^2, and σ_ε^2, as shown in (3.47) and (3.48). In fact for known θ, σ_γ^2, and σ_ε^2, the GQL estimating equation (3.74) in this linear model case yields the generalized least squares (GLS) estimator given by

$$\hat{\beta}_{GQL} \equiv \hat{\beta}_{GLS} = [\sum_{i=1}^{K} X_i' \Sigma_i^{-1} (\theta, \sigma_\gamma^2, \sigma_\varepsilon^2) X_i]^{-1} \sum_{i=1}^{K} X_i' \Sigma_i^{-1} (\theta, \sigma_\gamma^2, \sigma_\varepsilon^2) y_i, \qquad (3.75)$$

[see also (2.13)].

Note, however, that the consistent and efficient estimation of the scale parameters θ, σ_γ^2, and σ_ε^2 is not easy. Following Sutradhar (2004), we provide a second-order response based GQL estimating equation for these parameters and demonstrate that such GQL estimates are more efficient than the GMM based estimates discussed in Section 3.3.1. Also note that the IV based GMM approach [eqn. (3.42)] exploits the first-order response for the estimation of the correlation type scale parameter θ which produces a biased and inefficient estimate. See, for example, Sutradhar and Farrell (2007) for the effect of using first-order responses as opposed to second-order responses in estimating a similar (to θ) dynamic dependence parameter in the binary case. Further note that both the GMM and GQL exploit the moments of the data up to order four to construct the estimating equations for the scale parameters, but the ways of construction are completely different. This is because in the GMM approach moment functions for the first− and second-order responses are pooled together for all K independent individuals ignoring their variance and covariances, whereas in the GQL approach variance−covariance matrix based quasi-likelihood functions for each of the K independent individuals will be pooled together to construct the final GQL estimating equation.

Note that because K individuals are independent, it follows by applying the standard central limit theorem that the GQL/GLS estimator of β obtained from (3.75) asymptotically ($K \to \infty$) follows the multivariate normal distribution given as

$$\sqrt{K}(\hat{\beta}_{GQL} - \beta) \sim N\left(0, K\left[\sum_{i=1}^{K} X_i' \Sigma_i^{-1}(\theta, \sigma_\gamma^2, \sigma_\varepsilon^2) X_i\right]^{-1}\right). \quad (3.76)$$

b. GQL Estimation for $\xi = (\theta, \sigma_\gamma^2, \sigma_\varepsilon^2)'$

Note that under the present model (3.45), θ and σ_γ^2 are known to be dependence parameters, whereas σ_ε^2 is the variance of the error components of the model. By (3.47) and (3.48), these parameters, i.e., $\xi = (\theta, \sigma_\gamma^2, \sigma_\varepsilon^2)'$ are seen to be involved in the variances and covariances of the panel data. Thus, assuming that μ_{it} are known, we write an elementary sufficient statistic vector consisting of the corrected squares and the pairwise products of the responses, given by

$$\begin{aligned} s_i = [&(y_{i1} - \mu_{i1})^2, \ldots, (y_{it} - \mu_{it})^2, \ldots, (y_{iT} - \mu_{iT})^2, \\ &(y_{i1} - \mu_{i1})(y_{i2} - \mu_{i2}), \ldots, (y_{iu} - \mu_{iu})(y_{it} - \mu_{it}), \\ &\ldots, (y_{i,T-1} - \mu_{i,T-1})(y_{iT} - \mu_{iT})]' \end{aligned} \quad (3.77)$$

in order to construct a GQL estimating equation for $\xi = (\theta, \sigma_\gamma^2, \sigma_\varepsilon^2)'$. Following Sutradhar (2004), this GQL estimating equation has the form

$$\sum_{i=1}^{I} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is}^{-1}(s_i - \sigma_i) = 0, \quad (3.78)$$

where $\sigma_i = (\sigma_{i11}, \ldots, \sigma_{itt}, \ldots, \sigma_{iTT}, \sigma_{i12}, \ldots, \sigma_{iut}, \ldots, \sigma_{i,T-1,T})' = E(s_i)$, and $\Omega_{is} = \text{cov}(s_i)$.

Recall from (3.47) and (3.48) that

$$\sigma_{itt} = z_i^2 \sigma_\gamma^2 \left[\sum_{j=0}^{t-1} \theta^j\right]^2 + \sigma_\varepsilon^2 \sum_{j=0}^{t-1} \theta^{2j},$$

and

$$\sigma_{iut} = z_i^2 \sigma_\gamma^2 \sum_{j=0}^{t-1} \theta^j \sum_{k=0}^{u-1} \theta^k + \sigma_\varepsilon^2 \sum_{j=0}^{u-1} \theta^{t-u+2j}.$$

Furthermore the derivatives of σ_{itt} and σ_{iut}, with respect to θ, σ_γ^2, and σ_ε^2 are given in the equations from (3.60) to (3.63). Thus, $\partial \sigma_i' / \partial \xi$ for (3.78) is known. It is then clear that one may now solve (3.78) for $\xi = (\theta, \sigma_\gamma^2, \sigma_\in^2)'$, provided the covariance matrix of s_i, that is, $\Omega_{is} = \text{cov}(s_i)$ is known. Note, however, that as the distribution of $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ may not be known, it is then not possible to derive the true covariance matrix of s_i (3.77). But, as the consistency of the ξ parameter will not be affected by the choice of a 'working' matrix say $\Omega_{is,w}$ in place of Ω_{is}, we, for convenience, use a normal y_i based $\text{cov}(s_i) = \Omega_{is,N}$, say, and solve the 'working' GQL estimating equation

$$\sum_{i=1}^{I} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} (s_i - \sigma_i) = 0 \tag{3.79}$$

instead of (3.78).

Note that μ_i and Σ_i are the true mean vector and covariance matrix of the y_i, where y_i may or may not follow normal distribution. We now compute $\Omega_{is,N} = \text{cov}(s_i)$ under the 'working' assumption that

$$y_i \sim N(\mu_i, \Sigma_i). \tag{3.80}$$

As s_i contains corrected squares and pairwise products of the responses, the normality based fourth-order moments matrix $\Omega_{is,N}$ may be computed by using the general fourth-order moments given in (3.67), that is,

$$E[(Y_{i\ell} - \mu_{i\ell})(Y_{im} - \mu_{im})(Y_{iu} - \mu_{iu})(Y_{it} - \mu_{it})] = \sigma_{i\ell m}\sigma_{iut} + \sigma_{i\ell u}\sigma_{imt} + \sigma_{imu}\sigma_{i\ell t},$$

where σ_{itt} and σ_{iut} are the true variance and covariances and their formulas are given by (3.47) and (3.48), respectively.

Let $\hat{\xi}_{GQL} = (\hat{\theta}_{GQL}, \hat{\sigma}^2_{\gamma,GQL}, \hat{\sigma}^2_{\varepsilon,GQL})'$ be the solution of (3.79). Under some mild regularity conditions, it may be shown that asymptotically $(K \to \infty)$

$$K^{1/2}(\hat{\xi}_{GQL} - \xi) \sim N(0, KV^*_{GQL}), \tag{3.81}$$

where V^*_{GQL} is given by

$$V^*_{GQL} = \left[\sum_{i=1}^{K} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} \frac{\partial \sigma_i}{\partial \xi}\right]^{-1} \left[\sum_{i=1}^{K} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} \Omega_i \Omega_{is,N}^{-1} \frac{\partial \sigma_i'}{\partial \xi}\right] \times \left[\sum_{i=1}^{K} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} \frac{\partial \sigma_i}{\partial \xi}\right]^{-1}, \tag{3.82}$$

with Ω_{is} as the true covariance matrix of s_i, as in (3.78). Note that the asymptotic covariance matrix in (3.80) may be consistently estimated as

$$\hat{V}^*_{GQL} = \left[\sum_{i=1}^{K} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} \frac{\partial \sigma_i}{\partial \xi}\right]^{-1} \left[\sum_{i=1}^{K} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} (s_i - \sigma_i)(s_i - \sigma_i) \Omega_{is,N}^{-1} \frac{\partial \sigma_i'}{\partial \xi}\right]$$

$$\times \left[\sum_{i=1}^{K} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} \frac{\partial \sigma_i}{\partial \xi} \right]^{-1}. \quad (3.83)$$

Further note that if the true distributions of the model (3.45) errors are normal, then the asymptotic covariance matrix V_{GQL}^* in (3.82) reduces to

$$V_{GQL}^* = \left[\sum_{i=1}^{K} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} \frac{\partial \sigma_i}{\partial \xi} \right]^{-1}. \quad (3.84)$$

3.3.3 Asymptotic Efficiency Comparison

In this section, we compare the relative efficiency of the GQL approach to the GMM approach under normal errors. For this purpose, we compute the asymptotic variances of the GMM estimators of $\alpha = (\beta', \theta, \sigma_\gamma^2, \sigma_\varepsilon^2)'$ by

$$\mathrm{cov}(\hat{\alpha}_{GMM}) = \left\{ \frac{\partial \psi'}{\partial \alpha} C_N \frac{\partial \psi}{\partial \alpha'} \right\}^{-1} \quad (3.85)$$

[see eqn. (3.57)], and those of the GQL estimator of β and of $\xi = [\theta,\ \sigma_\gamma^2,\ \sigma_\varepsilon^2]'$ by

$$\mathrm{cov}(\hat{\beta}_{GQL}) = \left[\sum_{i=1}^{K} X_i' \Sigma_i^{-1} X_i \right]^{-1}, \text{ and } \mathrm{cov}(\hat{\xi}_{GQL}) = \left[\sum_{i=1}^{K} \frac{\partial \sigma_i'}{\partial \xi} \Omega_{is,N}^{-1} \frac{\partial \sigma_i}{\partial \xi} \right]^{-1} \quad (3.86)$$

[see (3.76) and (3.84)], respectively.

Now to reflect the asymptotic case, we consider $K = 500$ for the dynamic model (3.45). Furthermore, the panel data are usually collected over a small period of time, therefore we consider $T = 4$, for example. As far as the covariates are concerned, we choose two time-dependent covariates. The first covariate is considered to be:

$$x_{it1} = \begin{cases} 0 & \text{for } i = 1, \ldots, K/2;\ t = 1,2 \\ 1 & \text{for } i = 1, \ldots, K/2;\ t = 3,4 \\ 1 & \text{for } i = K/2+1, \ldots, K;\ t = 1, \ldots, 4, \end{cases} .$$

whereas the second covariate is chosen to be

$$x_{it2} = \begin{cases} 1 & \text{for } i = 1,\ldots,K/2;\ t = 1,2 \\ 1.5 & \text{for } i = 1,\ldots,K/2;\ t = 3,4 \\ 0 & \text{for } i = K/2+1,\ldots,K;\ t = 1,2 \\ 1 & \text{for} i = K/2+1,\ldots,K;\ t = 3,4. \end{cases}.$$

Next, we choose $\beta_1 = \beta_2 = 1.0$; $\theta = 0.3$ and 0.8, $\sigma_\gamma^2 = 0.5$, 0.8, and 1.2, and $\sigma_\varepsilon^2 = 1.0$. The diagonal elements (variances) of the covariance matrices from (3.85) and (3.86) are presented in Table 3.1 for the case when (a) $z_i = 1$, and in Table 3.2 when (b) $z_i \overset{iid}{\sim} N(0,1)$, for $i = 1,\ldots,500$.

Table 3.1 Comparison of asymptotic variances of the GQL and GMM estimators for the estimation of the regression parameters (β_1 and β_2), dynamic dependence parameter θ, and the variance components (σ_γ^2 and σ_ε^2), of a longitudinal mixed model for the normal panel data, with $T = 4$ and $K = 500$, $\beta_1 = \beta_2 = 1.0$, and $\sigma_\varepsilon^2 = 1.0$, and $z_i = 1$.

				Asymptotic Variances		
$z_i(i = 1,\ldots,500)$	θ	Method	Quantity	σ_γ^2= 0.5	0.8	1.2
1	0.3	GQL	$\text{Var}(\hat\beta_1)$	2.11×10.0^{-3}	2.18×10.0^{-3}	2.21×10.0^{-3}
			$\text{Var}(\hat\beta_2)$	1.57×10.0^{-3}	1.67×10.0^{-3}	1.74×10.0^{-3}
			$\text{Var}(\hat\theta)$	7.02×10.0^{-4}	3.78×10.0^{-4}	2.03×10.0^{-4}
			$\text{Var}(\hat\sigma_\gamma^2)$	2.31×10.0^{-2}	2.56×10.0^{-2}	2.90×10.0^{-2}
			$\text{Var}(\hat\sigma_\varepsilon^2)$	1.11×10.0^{-3}	1.20×10.0^{-3}	1.30×10.0^{-3}
		GMM	$\text{Var}(\hat\beta_1)$	1.93×10.0^{-3}	1.93×10.0^{-3}	1.94×10.0^{-3}
			$\text{Var}(\hat\beta_2)$	1.38×10.0^{-3}	1.38×10.0^{-3}	1.38×10.0^{-3}
			$\text{Var}(\hat\theta)$	2.40×10.0^{-3}	2.41×10.0^{-3}	2.41×10.0^{-3}
			$\text{Var}(\hat\sigma_\gamma^2)$	3450	3458	3471
			$\text{Var}(\hat\sigma_\varepsilon^2)$	3.04×10.0^{-3}	3.04×10.0^{-3}	3.05×10.0^{-3}
	0.8	GQL	$\text{Var}(\hat\beta_1)$	1.86×10.0^{-3}	1.43×10.0^{-3}	6.19×10.0^{-4}
			$\text{Var}(\hat\beta_2)$	1.77×10.0^{-3}	1.47×10.0^{-3}	7.28×10.0^{-4}
			$\text{Var}(\hat\theta)$	1.20×10.0^{-4}	3.72×10.0^{-5}	6.96×10.0^{-6}
			$\text{Var}(\hat\sigma_\gamma^2)$	3.35×10.0^{-3}	4.38×10.0^{-3}	4.34×10.0^{-3}
			$\text{Var}(\hat\sigma_\varepsilon^2)$	1.76×10.0^{-3}	2.13×10.0^{-3}	2.20×10.0^{-3}
		GMM	$\text{Var}(\hat\beta_1)$	4.08×10.0^{-3}	4.09×10.0^{-3}	4.10×10.0^{-3}
			$\text{Var}(\hat\beta_2)$	3.09×10.0^{-3}	3.09×10.0^{-3}	3.09×10.0^{-3}
			$\text{Var}(\hat\theta)$	0.132	0.133	0.135
			$\text{Var}(\hat\sigma_\gamma^2)$	31,457	31,762	32,250
			$\text{Var}(\hat\sigma_\varepsilon^2)$	1.93×10.0^{-2}	1.98×10.0^{-2}	2.04×10.0^{-2}

Note that when $z_i = 1$ in (3.45), this reduces to the standard dynamic mixed model. For this standard case, it is clear from Table 3.1 that the GMM approach produces regression estimates with variances the same as, or smaller than the GQL approach only when the dynamic dependence parameter is small ($\theta = 0.3$). With regard to the estimation of the other parameters including the estimation of the dynamic dependence parameter, the GQL approach produces estimates with smaller variances than the GMM estimates. In fact, the GMM approach cannot be trusted

Table 3.2 Comparison of asymptotic variances of the GQL and GMM estimators for the estimation of the regression parameters (β_1 and β_2), dynamic dependence parameter θ, and the variance components (σ^2_γ and σ^2_ε), of a longitudinal mixed model for the normal panel data, with $T = 4$ and $K = 500$, $\beta_1 = \beta_2 = 1.0$, and $\sigma^2_\varepsilon = 1.0$, and $z_i \overset{\text{iid}}{\sim} N(0,1)$.

				Asymptotic Variances		
$z_i (i = 1,\ldots,500)$	θ	Method	Quantity	$\sigma^2_\gamma = 0.5$	0.8	1.2
$N(0,1)$	0.3	GQL	$\text{Var}(\hat{\beta}_1)$	1.85×10.0^{-3}	1.77×10.0^{-3}	1.97×10.0^{-3}
			$\text{Var}(\hat{\beta}_2)$	1.38×10.0^{-3}	1.34×10.0^{-3}	1.49×10.0^{-3}
			$\text{Var}(\hat{\theta})$	1.27×10.0^{-4}	7.50×10.0^{-5}	4.84×10.0^{-5}
			$\text{Var}(\hat{\sigma}^2_\gamma)$	5.62×10.0^{-4}	1.29×10.0^{-3}	2.57×10.0^{-3}
			$\text{Var}(\hat{\sigma}^2_\varepsilon)$	1.05×10.0^{-3}	1.07×10.0^{-3}	1.10×10.0^{-3}
		GMM	$\text{Var}(\hat{\beta}_1)$	2.52×10.0^{-3}	2.88×10.0^{-3}	3.36×10.0^{-3}
			$\text{Var}(\hat{\beta}_2)$	1.85×10.0^{-3}	2.14×10.0^{-3}	2.53×10.0^{-3}
			$\text{Var}(\hat{\theta})$	2.53×10.0^{-3}	2.68×10.0^{-3}	2.96×10.0^{-3}
			$\text{Var}(\hat{\sigma}^2_\gamma)$	5.27×10.0^{-2}	8.82×10.0^{-2}	0.156
			$\text{Var}(\hat{\sigma}^2_\varepsilon)$	9.75×10.0^{-3}	1.99×10.0^{-2}	4.28×10.0^{-2}
	0.8	GQL	$\text{Var}(\hat{\beta}_1)$	2.44×10.0^{-3}	1.69×10.0^{-3}	2.87×10.0^{-3}
			$\text{Var}(\hat{\beta}_2)$	4.28×10.0^{-3}	3.61×10.0^{-3}	2.03×10.0^{-3}
			$\text{Var}(\hat{\theta})$	5.62×10.0^{-7}	3.63×10.0^{-8}	8.17×10.0^{-7}
			$\text{Var}(\hat{\sigma}^2_\gamma)$	2.66×10.0^{-4}	6.49×10.0^{-4}	1.46×10.0^{-3}
			$\text{Var}(\hat{\sigma}^2_\varepsilon)$	1.00×10.0^{-3}	1.00×10.0^{-3}	1.00×10.0^{-3}
		GMM	$\text{Var}(\hat{\beta}_1)$	6.82×10.0^{-3}	8.48×10.0^{-3}	1.07×10.0^{-2}
			$\text{Var}(\hat{\beta}_2)$	5.53×10.0^{-3}	6.99×10.0^{-3}	8.94×10.0^{-3}
			$\text{Var}(\hat{\theta})$	0.799	0.305	0.192
			$\text{Var}(\hat{\sigma}^2_\gamma)$	1.108	0.521	0.422
			$\text{Var}(\hat{\sigma}^2_\varepsilon)$	9.602	8.567	11.421

for the estimation of the variance component (σ^2_γ) of the random effects. This is because this approach produces σ^2_γ estimates with huge variances such as 3450 when the true $\sigma^2_\gamma = 0.5$, whereas the corresponding variance under the GQL approach is only 0.0231. For large dependence parameter $\theta = 0.8$, the GQL approach uniformly produces estimates for all parameters including the regression effects with smaller variances than the GMM approach. This asymptotic comparison between the GQL and GMM approaches indicates that in general the GQL approach is much more efficient than the GMM approach in estimating the parameters of the dynamic dependence model (3.45).

We have further considered a less realistic situation where the model (3.45) now allows individual fixed covariates z_i with random influence γ_i. For example, in this asymptotic empirical study, we generated the 500 values of z_i from $N(0,1)$. The asymptotic variances of the estimators for all five parameters β_1, β_2, θ, σ^2_γ, and σ^2_ε, under the GMM and GQL approaches are now reported in Table 3.2. When compared with Table 3.1, it is clear that in this nonstandard case, the GMM approach improves in estimating the variance component σ^2_γ. The GQL approach, however, is uniformly better than the GMM approach in estimating all parameters including the regression effects, irrespective of the situations whether the panel data have small or large dynamic dependence. For example, when $\theta = 0.3$ and $\sigma_\gamma^2 = 0.8$, the GQL estimates of β_1 and β_2 are, respectively,

$$\frac{2.88 \times 10.0^{-3}}{1.77 \times 10.0^{-3}} = 1.63 \quad \text{and} \quad \frac{2.14 \times 10.0^{-3}}{1.34 \times 10.0^{-3}} = 1.60$$

times more efficient than the corresponding GMM estimates. For the estimation of θ, σ_γ^2, and σ_ε^2, the GQL approach appears to outperform the GMM approach. For example, for the same set of parameter values, the GQL estimates of θ, σ_γ^2, and σ_ε^2 are, respectively,

$$\frac{2.68 \times 10.0^{-3}}{7.50 \times 10.0^{-5}} = 35.73, \quad \frac{0.0882}{1.29 \times 10.0^{-3}} = 68.37, \;\text{and}\; \frac{0.0199}{1.07 \times 10.0^{-3}} = 18.60$$

times more efficient than the corresponding GMM estimates. For the larger dynamic dependence parameter $\theta = 0.8$, the GMM performs much worse as compared to the GQL approach. In summary, the GQL approach performs much better than the GMM approach in estimating all five parameters, its performance being extraordinarily better in estimating the dynamic dependence parameter θ, and variance components σ_γ^2 and σ_ε^2.

Exercises

3.1. (Section 3.1.1) [Inverse of the covariance matrix under linear mixed model] Let C be a nonsingular matrix of dimension $T \times T$, and U and S be two T-dimensional column vectors. Then

$$[C + US']^{-1} = C^{-1} - \frac{1}{1 + S'C^{-1}U}(C^{-1}U)(S'C^{-1}).$$

This result immediately gives the inverse of $\Sigma_i = \sigma_\varepsilon^2 C_i + \sigma_\gamma^2 1_{T_i} 1'_{T_i}$ as in (3.11).

3.2. (Section 3.2.2. a) [LSDV estimators for a simpler linear dynamic mixed model] For scalar β, the formulas for the LSDV estimators of θ and β given by (3.32) may be simplified as

$$\hat{\theta}_{lsdv} = \frac{\sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}^2 \sum_{i=1}^K \sum_{t=2}^T \tilde{y}_{it}\tilde{y}_{i,t-1} - \sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}\tilde{y}_{i,t-1} \sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}\tilde{y}_{it}}{\sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}^2 \sum_{i=1}^K \sum_{t=2}^T \tilde{y}_{i,t-1}^2 - [\sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}\tilde{y}_{i,t-1}]^2}$$

and

$$\hat{\beta}_{lsdv} = \frac{-\sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}\tilde{y}_{i,t-1} \sum_{i=1}^K \sum_{t=2}^T \tilde{y}_{it}\tilde{y}_{i,t-1} + \sum_{i=1}^K \sum_{t=2}^T \tilde{y}_{i,t-1}^2 \sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}\tilde{y}_{it}}{\sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}^2 \sum_{i=1}^K \sum_{t=2}^T \tilde{y}_{i,t-1}^2 - [\sum_{i=1}^K \sum_{t=2}^T \tilde{w}_{it}\tilde{y}_{i,t-1}]^2},$$

[Bun and Carree (2005, eqns. (3) and (4))] respectively.

3.3. (Section 3.2.2. a) [Bias-corrected LSDV estimators for special cases]
For scalar β, the BCLSDV estimator for β has the same formula

$$\hat{\beta}_{bclsdv} = \hat{\beta}_{lsdv} + \hat{\xi}(\hat{\theta}_{lsdv} - \hat{\theta}_{bclsdv})$$

as in (3.33), whereas for $T = 3$ and 4, the BCLSDV estimator of θ has explicit formulas given by

$$\hat{\theta}_{bclsdv} = \hat{\theta}_{lsdv} + \frac{\hat{\sigma}_\varepsilon^2}{2(1-\hat{\rho}_{wy_{-1}}^2)\hat{\sigma}_{y_{-1}}^2}$$

and

$$\hat{\theta}_{bclsdv} = \frac{6\hat{\theta}_{lsdv} + 2\hat{\sigma}_\varepsilon^2/[(1-\hat{\rho}_{wy_{-1}}^2)\hat{\sigma}_{y_{-1}}^2]}{6 - \hat{\sigma}_\varepsilon^2/[(1-\hat{\rho}_{wy_{-1}}^2)\hat{\sigma}_{y_{-1}}^2]}$$

[Bun and Carree (2005, eqns. (14) and (15))] respectively.

3.4. (Section 3.2.2. b) [Instrumental variable estimators]
Demonstrate that the formula for the IV estimators for θ and β given by (3.37) is the same [see also Hsiao (2003, eqn. (4.3.32))] as

$$\begin{bmatrix} \hat{\theta}_{iv} \\ \hat{\beta}_{iv} \end{bmatrix} = \left[\sum_{i=1}^{K}\sum_{t=4}^{T} \begin{pmatrix} (y_{i,t-2}^*)(y_{i,t-1}^*) & (y_{i,t-2}^*)(w_{it} - w_{i,t-1})' \\ (w_{it} - w_{i,t-1})(y_{i,t-1}^*) & (w_{it} - w_{i,t-1})(w_{it} - w_{i,t-1})' \end{pmatrix}\right]^{-1} \times \left[\sum_{i=1}^{K}\sum_{t=4}^{T} \begin{pmatrix} y_{i,t-2}^* \\ w_{it} - w_{i,t-1} \end{pmatrix} (y_{it}^*)\right], \tag{3.87}$$

where $y_{i,t-1}^* = (y_{i,t-1} - y_{i,t-2})$, for example.

3.5. (Sections 3.3.1-3.3.2) [GMM and GQL approaches with independent t error distributions]
Suppose that the errors ε_{it} in the model (3.45) are now independently distributed (id) [as opposed to independently and identically distributed (iid)] as

$$\varepsilon_{it} \overset{\text{id}}{\sim} (0, \lambda_t^2\sigma_\varepsilon^2),$$

where $\lambda_1, \ldots, \lambda_t, \ldots, \lambda_T$ are random and independent scales each with a one-parameter (ν) based inverted gamma distribution given as

$$g(\lambda_t) = \frac{2(\nu/2)^{-1/2}}{\Gamma(\nu/2)} \exp\left\{-\frac{1}{2}(\nu/\lambda_t^2)\right\}\left[\frac{\nu}{2\lambda_t^2}\right]^{(\nu+1)/2}, \ \lambda_t > 0,$$

yielding the t-distribution for ε_{it} as

$$f(\varepsilon_{it}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{1/2}\nu^{1/2}\Gamma\left(\frac{\nu}{2}\right)}\sigma_{\varepsilon}^{-1}\left[1+\frac{\{\varepsilon_{it}/\sigma_{\varepsilon}\}^2}{\nu}\right]^{-(\nu+1)/2},$$

[see Sutradhar (1988, p. 176), e.g.] with $\nu > 0$ degrees of freedom.
(a) Use the density function of λ_t and show that $E[\lambda_t^2] = \nu/(\nu-2)$. Then show that the unconditional variance of ε_{it}; that is,

$$\text{var}[\varepsilon_{it}] = E_{\lambda_t}\text{var}[\varepsilon_{it}|\lambda_t] + \text{var}_{\lambda_t}E[\varepsilon_{it}|\lambda_t] = \frac{\nu}{\nu-2}\sigma_{\varepsilon}^2.$$

(b) By using the t-density of ε_{it}, it may also be shown directly that $\text{var}[\varepsilon_{it}] = [\nu/(\nu-2)]\sigma_{\varepsilon}^2$.
(c) Now demonstrate that for known $\nu > 2$, the 'working' normality (N) based GMM and GQL estimation given in Sections 3.3.1 and 3.3.2 can be carried out simply by replacing the formulas for the variances (σ_{itt}) (3.47) and covariances (σ_{iut}) (3.48) with

$$\text{var}[Y_{it}] = \sigma_{itt} = z_i^2\sigma_{\gamma}^2\left\{\sum_{j=0}^{t-1}\theta^j\right\}^2 + \frac{\nu}{\nu-2}\sigma_{\varepsilon}^2\sum_{j=0}^{t-1}\theta^{2j}, \tag{3.88}$$

and

$$\text{cov}[Y_{iu}, Y_{it}] = \sigma_{iut} = z_i^2\sigma_{\gamma}^2\sum_{j=0}^{t-1}\theta^j\sum_{k=0}^{u-1}\theta^k + \frac{\nu}{\nu-2}\sigma_{\varepsilon}^2\sum_{j=0}^{u-1}\theta^{t-u+2j}, \tag{3.89}$$

respectively.

References

1. Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
2. Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Rev. Eco. Statist.,* 58, 277 − −298.
3. Balestra, P. & Nerlove, M. (1966). Pooling cross-section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica,* 34, 585 − 612.
4. Bun, M. J. G. & Carree, M. A. (2005). Bias-corrected estimation in dynamic panel data models. *J. Business & Econo. Statist.,* 23, 200 − 210.
5. Chow, S. C. & Shao, J. (1988). A new procedure for the estimation of variance components. *Statist. Probab. Lett.,* 6, 349 − 355.
6. Diggle, P. J., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford Science, Oxford: Clarendon Press.
7. Hansen, L.-P. (1982). Large sample properties of generalized method of moment estimators. *Econometrica,* 50, 1029 − 1054.
8. Herbach, L. H. (1959). Properties of type II analysis of variance test. *Ann. Math. Statist.,* 30, 939 − 959.
9. Hsiao, C. (2003). *Analysis of Panel Data.* Cambridge, U.K.: University Press.
10. Jiang, J. & Zhang, W. (2001). Robust estimation in generalized linear mixed models. *Biometrika,* 88, 753 − 765.

11. LaMotte, L. R. (1973). On non-negative quadratic unbiased estimation of variance components. *J. Amer. Statist. Assoc.,* 68, 728 – 730.
12. McCullagh, P. (1983). Qusilikelihood functions. *Ann. Statist.,* 11, 59-67.
13. Mathew, T., Sinha, B. K., & Sutradhar, B. C. (1992). Nonnegative estimation of variance components in unbalanced mixed models with two variance components. *J. Multivariate Anal.,* 42, 77 – 101.
14. Rao, R. P., Sutradhar, B. C., & Pandit, V. N. (2010). GMM versus GQL inferences in linear dynamic panel data models. *Braz. J. Probab. Statist.,* to appear.
15. Searle, S. R. (1971). *Linear Models.* New York: John Wiley & Sons.
16. Sneddon, G. & Sutradhar, B. C. (2004). On semi-parametric familial-longitudinal models. *Statist. Probab. Lett.,* 69, 369 – 379.
17. Sutradhar, B. C. (1988). Testing linear hypothesis with t error variable. *Sankhya B: Indian J. Statist.,* 50, 175 – 180.
18. Sutradhar, B. C. (1997). A multivariate approach for estimating the random effects variance component in one-way random effects model. *Stat. Probab. Lett.,* 33, 333 – 339.
19. Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statist. Sci.,* 18, 377 – 393.
20. Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhya B: Indian J. of Statist.,* 66, 261 – 289.
21. Sutradhar, B. C. & Farrell, P. J. (2007). On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data. *Sankhya B: Indian J. Statist.,* 69, 448 – 467.
22. Thompson, W. A. Jr. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.,* 33, 273 – 289.
23. Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York: Springer.
24. Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika,* 61, 439 – 447.

Chapter 4
Familial Models for Count Data

Familial models for count data are also known as Poisson mixed models for count data. In this setup, count responses along with a set of multidimensional covariates are collected from the members of a large number of independent families. Let y_{ij} denote the count response for the jth $(j = 1, \ldots, n_i)$ member on the ith $(i = 1, \ldots, K)$ family/cluster. Also, let $x_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ denote the p covariates associated with the count response y_{ij}. For example, in a health economics study, a state government may be interested to know the effects of certain socioeconomic and epidemiological covariates such as gender, education level, and age on the number of visits by a family member to the house physician in a particular year. Note that in this problem it is also likely that the count responses of the members of a family are influenced by a common random family effect, say γ_i. This makes the count responses of any two members of the same family correlated, and this correlation is usually referred to as the familial correlation. It is of scientific interest to find the effects of the covariates on the count responses of an individual member after taking the familial correlations into account.

In Section 4.1, we provide the marginal (unconditional) distributional properties of the count response variable y_{ij} as well as the unconditional familial correlation structure under suitable distributional assumptions for the random effects. Frequently, it is assumed that the random effects follow normal distributions [Breslow and Clayton (1993); Lee and Nelder (1996)]. One of the main reasons for this assumption is that the familial Poisson mixed models or generalized linear mixed models (GLMMs) in general are generated from the well-known generalized linear models (GLMs) [McCullagh and Nelder (1989)] by adding random effects to the linear predictor. Under this normality assumption for the random effects, in Section 4.2, various inference techniques such as the method of moments, likelihood approximations, and quasi-likelihood approaches are discussed for the estimation of the effects of the covariates and the familial correlation index parameter.

Note that in some situations, the responses of the family members may be influenced by more than one common random family effect. If this happens, it is also important to recognize that these multidimensional random effects may play different roles in different setup. For example, in case of two random effects, some authors

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_4,

such as Lin (1997), Jiang (1998), and Sutradhar and Rao (2003) have assumed that these random effects follow a two-factor factorial design or a nested design. There are, however, other situations in practice, where each of the count responses for a given family is influenced by two distinct random effects with two different components of dispersion. See, for example, Jowaheer, Sutradhar, and Sneddon (2009). In Section 4.3, we accommodate the different natures of the random effects and discuss in detail the inferences in Poisson mixed models with two variance components. As far as the distributional assumptions are concerned, similar to Section 4.2, it is assumed that the random effects follow a normal distribution. A Poisson mixed model with more than two random effects may similarly be studied, but the inferences for this type of complex model are not discussed in detail as they rarely arise in practice.

In Section 4.4, this distributional assumption is relaxed and an alternative inference technique, namely a semiparametric approach is discussed. In Section 4.5, a Monte Carlo (MC) based likelihood estimation approach is outlined. The drawbacks of these general approaches are also pointed out.

4.1 Poisson Mixed Models and Basic Properties

Let $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$ be the $n_i \times 1$ vector of count responses from n_i members of the ith $(i = 1, \ldots, K)$ family. Let β be a $p \times 1$ vector of unknown fixed effects of x_{ij} on y_{ij}, x_{ij} being the p-dimensional covariate vector for the jth $(j = 1, \ldots, n_i)$ member of the ith family. Suppose that conditional on the random family effect γ_i, n_i counts due to the ith family are independent. The data of this type can be modelled as

$$f(y_i|\gamma_i) = \frac{1}{\Pi_{j=1}^{n_i} y_{ij}!} \exp\left(\sum_{j=1}^{n_i} y_{ij}\eta_{ij} - \sum_{j=1}^{n_i} \exp(\eta_{ij})\right), \tag{4.1}$$

where $f(y_i|\gamma_i)$ denotes the conditional probability density of y_i for a given γ_i, and where η_{ij} is a linear predictor defined as $\eta_{ij} = x'_{ij}\beta + \gamma_i$. Further suppose that γ_i has an unspecified distribution with mean 0 and variance σ_γ^2 and γ_i are independent, that is, $\gamma_i \overset{\text{iid}}{\sim} (0, \sigma_\gamma^2)$. For $\gamma_i^* = \gamma_i/\sigma_\gamma$, the linear predictor in (4.1) may then be expressed as

$$\eta_{ij}(\gamma_i) = x'_{ij}\beta + \gamma_i = x'_{ij}\beta + \sigma_\gamma\gamma_i^*, \tag{4.2}$$

where $\gamma_i^* \overset{\text{iid}}{\sim} (0, 1)$. Note that as shown in the following lemma, the variance component of the random effects (σ_γ^2) indicates the possible overdispersion in the Poisson count data. This is why this scale parameter is often referred to as an overdispersion index parameter. Lemma 4.1 also shows that σ_γ^2 plays a role of a familial correlation parameter. This is because when $\sigma_\gamma^2 = 0$ the count responses of the family members become independent. For this reason, one may refer to this σ_γ^2 parameter as a familial or structural correlation index parameter. In practice, it is of interest to estimate both regression parameter vector β and the overdispersion or familial correlation index parameter σ_γ^2, as consistently and efficiently as possible.

For convenience, for the development of the estimation techniques for β and σ_γ^2 in the followup sections, we first provide the conditional (on γ_i^*) as well as the unconditional first− and second-order moments of the count variables

$$Y_{i1},\ldots,Y_{ij},\ldots,Y_{in_i},$$

in Lemma 4.1.

Lemma 4.1. Conditional on γ_i^*, the means and the variances of Y_{ij}, and the pairwise covariances between Y_{ij} and Y_{ik} for $j \neq k,\ j,\ k = 1,\ldots,n_i$ are given by

$$E[Y_{ij}|\gamma_i^*] = \text{var}[Y_{ij}|\gamma_i^*] = \mu_{ij}^* = \exp(x_{ij}'\beta + \sigma_\gamma\gamma_i^*) \tag{4.3}$$

$$\text{cov}[(Y_{ij},Y_{i,k})|\gamma_i^*] = 0, \tag{4.4}$$

and for $\gamma_i^* \overset{\text{iid}}{\sim} N(0,1)$, the corresponding unconditional means, variances, and covariances are given by

$$E[Y_{ij}] = \mu_{ij} = \exp(x_{ij}'\beta + \frac{1}{2}\sigma_\gamma^2) \tag{4.5}$$

$$\text{var}[Y_{ij}] = \mu_{ij} + [\exp(\sigma_\gamma^2) - 1]\mu_{ij}^2 \tag{4.6}$$

$$\text{cov}[Y_{ij},Y_{ik}] = \mu_{ij}\mu_{ik}[\exp(\sigma_\gamma^2) - 1], \tag{4.7}$$

yielding the pairwise familial correlations as

$$\text{corr}[Y_{ij},Y_{ik}] = \frac{\exp(\sigma_\gamma^2) - 1}{[\{\mu_{ij}^{-1} + (\exp(\sigma_\gamma^2) - 1)\}\{\mu_{ik}^{-1} + (\exp(\sigma_\gamma^2) - 1)\}]^{1/2}}. \tag{4.8}$$

Proof: For an auxiliary parameter s, it follows from (4.1) that the moment generating function (mgf) of Y_{ij} conditional on γ_i^*, is given by

$$M_{Y_{ij}|\gamma_i^*}(s) = E[\exp(sY_{ij})] = \exp[\mu_{ij}^*\{\exp(s) - 1\}], \tag{4.9}$$

where $\mu_{ij}^* = \exp(\eta_{ij}) = \exp(x_{ij}'\beta + \sigma_\gamma\gamma_i^*)$.

Now for a positive integer r, by evaluating the rth order derivative of the mgf in (4.9) with respect to s, at $s = 0$, that is, by simplifying

$$\frac{\partial^r}{\partial s^r} M_{Y_{ij}|\gamma_i^*}(s)|_{s=0},$$

one obtains $E[Y_{it}^r|\gamma_i^*]$. For the special cases with $r = 1,2$, one obtains

$$E[Y_{ij}|\gamma_i^*] = \mu_{ij}^*$$

$$E[Y_{ij}^2|\gamma_i^*] = \mu_{ij}^* + \mu_{ij}^{*\,2}, \tag{4.10}$$

and similarly for $r = 3,4$, the conditional moments have the formulas:

$$\begin{aligned} E[Y_{ij}^3|\gamma_i^*] &= \mu_{ij}^* + 3\mu_{ij}^{*2} + \mu_{ij}^{*3} \\ E[Y_{ij}^4|\gamma_i^*] &= \mu_{ij}^* + 7\mu_{ij}^{*2} + 6\mu_{ij}^{*3} + \mu_{ij}^{*4}. \end{aligned} \tag{4.11}$$

The results in (4.3) follow from (4.10). The result in (4.4) follows from the fact [see also (4.1)] that conditional on γ_i^*, the count responses of any two members of a family are independent, implying that

$$E[Y_{ij}Y_{ik}|\gamma_i^*] = E[Y_{ij}|\gamma_i^*]E[Y_{ik}|\gamma_i^*] = \mu_{ij}^*\mu_{ik}^*, \tag{4.12}$$

for $j \neq k,\ j,k = 1,\ldots,n_i$. $\diamond$

Next, under the assumption that $\gamma_i^* \stackrel{\text{iid}}{\sim} N(0,1)$, for an auxiliary parameter s, one writes the moment generating function of γ_i^* as

$$E[\exp(s\gamma_i^*)] = \exp(\frac{1}{2}s^2). \tag{4.13}$$

This moment generating function along with the formula for μ_{ij}^*, that is, $\mu_{ij}^* = \exp(x_{ij}'\beta)[\exp(\sigma_\gamma\gamma_i^*)]$, can be used to derive the unconditional first$-$ and the second-order moments as

$$E[Y_{ij}] = E_{\gamma_i^*}E[Y_{ij}|\gamma_i^*] = E_{\gamma_i^*}[\mu_{ij}^*] = \exp(x_{ij}'\beta + \frac{1}{2}\sigma_\gamma^2) = \mu_{ij} \tag{4.14}$$

$$E[Y_{ij}^2] = E_{\gamma_i^*}E[Y_{ij}^2|\gamma_i^*] = E_{\gamma_i^*}[\mu_{ij}^* + \mu_{ij}^{*2}] = \mu_{ij} + \exp(\sigma_\gamma^2)\mu_{ij}^2 \tag{4.15}$$

$$E[Y_{ij}Y_{ik}] = E_{\gamma_i^*}E[Y_{ij}Y_{ik}|\gamma_i^*] = E_{\gamma_i}[\mu_{ij}^*\mu_{ik}^*] = \exp(\sigma_\gamma^2)\mu_{ij}\mu_{ik}, \tag{4.16}$$

yielding the unconditional means, variances, and covariances as in $(4.5)-(4.7)$.

Note that as mentioned earlier, it is clear from (4.6) and (4.8) that σ_γ^2 may be referred to as the overdispersion or familial correlation index parameter. This is because, the equation (4.6), for example, indicates that σ_γ^2 plays an important role in understanding the dispersion of the response variable Y_{ij}. To be specific, when $\sigma_\gamma^2 = 0$, Y_{ij} has the same dispersion as that of the Poisson data, whereas a slight increase in the value of σ_γ^2 may cause very large dispersion in the data, especially when $\exp(x_{ij}'\beta)$ is large.

Further note that in order to understand the model or the properties such as the model based mean and the variance of the data, it is of interest to estimate both regression effects β and the variance component of the random effects, σ_γ^2. Various techniques for the estimation of these parameters are discussed in the next section under a fully parametric model setup. Estimation in a similar parametric model setup but with multiple random effects, is discussed in Section 4.3. In Section 4.4, we provide the estimation techniques for the regression and single variance component parameters in a semiparametric setup, where no assumptions are made for the distributions of the random effects, instead it is assumed that the moments up to order four are known. In Section 4.5, we outline the estimation of the parameters in a nonparametric setup.

4.2 Estimation for Single Random Effect Based Parametric Mixed Models

In this parametric model setup, it is assumed that γ_i^* follows a specified distribution such as

$$\gamma_i^* \overset{\text{iid}}{\sim} N(0,1). \tag{4.17}$$

Also, it is assumed that conditional on γ_i^*, the count responses of the members of the ith family are independent, and marginally they follow a Poisson density leading to the likelihood function in (4.1).

4.2.1 Exact Likelihood Estimation and Drawbacks

The log-likelihood function based on (4.1) – (4.2) and (4.17) is given by

$$\log L(\beta,\sigma_\gamma) = -\sum_{i=1}^{K}\sum_{j=1}^{n_i} \log y_{ij}! + \sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}x'_{ij}\beta + \sum_{i=1}^{K} \log J_i, \tag{4.18}$$

with

$$J_i = \int_{-\infty}^{\infty} \exp[s_i(\gamma_i^*)] g_N(\gamma_i^*|1) d\gamma_i^*, \tag{4.19}$$

where $g_N(\gamma_i^*|1)$ is the standard normal density of γ_i^* and

$$s_i(\gamma_i^*) = \sigma_\gamma \gamma_i^* \sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} \exp(x'_{ij}\beta + \sigma_\gamma \gamma_i^*). \tag{4.20}$$

Note that the exact computation of the above integral J_i is not possible. The NLMIXED procedure in SAS, for example, uses a numerical approximation to this integral. Some authors use a simulation approach or binomial approximation to evaluate this integral. Fahrmeir and Tutz (1994, Chapter 7) [see also Jiang (1998)], for example, use a simulation technique to evaluate such integrals. More specifically, in the simulation technique, for a large N such as $N = 1000$, J_i is replaced by a simulation based $J_i^{(s)}$, where

$$J_i^{(s)} = \frac{1}{N} \sum_{w=1}^{N} \exp[s_i(\gamma_{iw}^*)], \tag{4.21}$$

where γ_{iw}^* is a sequence of standard normal values for $w = 1,\ldots,N$. Ten Have and Morabia [1999, eqn. (7)], for example, have used the standardized binomial approximation to evaluate similar integrals. For a known reasonably big V such as $V = 5$, let $v_i \sim \text{binomial}(V, 1/2)$. Because γ_i^* has the standard normal distribution, consider

$$\gamma_i^* = \frac{v_i - V(1/2)}{V(1/2)(1/2)}.$$

One may then approximate the integral J_i by a binomial approximation based integral $J_i^{(b)}$ defined as

$$J_i^{(b)} = \sum_{v_i=0}^{V} \exp[s_i(v_i)] \binom{V}{v_i} (1/2)^{v_i}(1/2)^{V-v_i}, \tag{4.22}$$

where

$$\exp[s_i(v_i)] = [\exp\{s_i(\gamma_i^*)\}]_{[\gamma_i^*=\{v_i-V(1/2)\}/\{V(1/2)(1/2)\}]},$$

with $s_i(\gamma_i^*)$ as in (20).

Next, by using, for example, $J_i^{(s)}$ for J_i in (4.18), one solves the simulation based approximate likelihood estimating equations

$$U_1^{(s)}(\beta, \sigma_\gamma^2) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}[y_{ij} - \frac{A_i^{(s)}}{J_i^{(s)}}]x_{ij} = 0, \tag{4.23}$$

and

$$U_2^{(s)}(\beta, \sigma_\gamma^2) = \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{n_i}[\sigma_\gamma^{-1}\frac{M_i^{(s)}}{J_i^{(s)}}] = 0, \tag{4.24}$$

for β and σ_γ^2, respectively. In (4.23) – (4.24), $A_i^{(s)}$ and $M_i^{(s)}$ are given by

$$\begin{aligned} A_i^{(s)} &= \frac{1}{N}\sum_{w=1}^{N} \exp[s_i(\gamma_{iw}^*)]\exp[\eta_{ij}(\gamma_{iw}^*)], \\ M_i^{(s)} &= \frac{1}{N}\sum_{w=1}^{N} \exp[s_i(\gamma_{iw}^*)]\left[\gamma_{iw}^*\sum_{j=1}^{n_i}\{y_{ij} - \exp[\eta_{ij}(\gamma_{iw}^*)]\}\right]. \end{aligned} \tag{4.25}$$

Some Drawbacks

(1) One of the main difficulties of the likelihood estimation approach is the complexity involved in computing the Fisher information matrix (more reliable than using the Hessian matrix) for the purpose of computing the standard errors of the likelihood estimates of β and σ_γ^2. This is evident, for example, from the formula for the expectation of the second derivative of the log likelihood function with respect to β, given by

$$E[\frac{\partial U_1^{(s)}(\beta,\sigma_\gamma^2)}{\partial \beta'}] = \sum_{y_{i1}=0}^{\infty} \cdots \sum_{y_{in_i}=0}^{\infty} [\{M_i^{(s)}\}^2/J_i^{(s)}] \exp\left[\sum_{j=1}^{n_i} y_{ij}x'_{ij}\beta\right] / \Pi_{j=1}^{n_i} y_{ij}!, \tag{4.26}$$

which is computationally quite involved, in particular for large n_i.

(2) Computations become cumbersome when the mixed model involves multi-dimensional random effects [e.g. Jiang (1998)].

(3) The likelihood approach naturally would be of no use for the inferences in the extended familial longitudinal model, where further responses are collected over a period of time from the members of all families. This is because it is either impossible or extremely complex to write a likelihood function for the repeated count responses; they are being longitudinally correlated conditional on the random effects.

To avoid the above and other possible difficulties with the exact likelihood approach, some authors such as Breslow and Clayton (1993) have suggested a penalized quasi-likelihood (PQL) approximation; Lee and Nelder (1996) have used a hierarchial likelihood (HL) approximation. These approaches estimate β and σ_γ^2 through the estimation/prediction of the random effects γ_i. Breslow and Lin (1995) have, however, cautioned in the context of a binary mixed model that the PQL approach yields consistent estimates for both β and σ_γ^2 provided the true σ_γ^2 value is small ($\leq$ 0.5). Sutradhar and Qu (1998) have demonstrated this inconsistency problem under a Poisson mixed model, and further proposed a small σ_γ^2 asymptotic approach to develop a likelihood approximation (LA) that produces less-biased estimates than the PQL approach even if true σ_γ^2 is more than 0.5. For the same Poisson mixed model, Chowdhury and Sutradhar (2009) have shown that the HL approach of Lee and Nelder (1996) also suffers from similar bias or inconsistency problems. In the following three subsections, these PQL, LA, and HL approaches are discussed in brief for the sake of completeness.

4.2.2 Penalized Quasi-Likelihood Approach

For the Poisson mixed model defined through (4.1) – (4.2) and (4.17), the log of the quasi-likelihood function derived by Breslow and Clayton [1993, eqn. (5), p. 11] reduces to

$$ql(\beta,\sigma_\gamma^2,\tilde{\gamma}) = -\frac{1}{2}\sum_{i=1}^{K} \log\left(1+\sigma_\gamma^2 \sum_{j=1}^{n_i} \exp(x'_{ij}\beta+\tilde{\gamma}_i)\right) - \sum_{i=1}^{K} h(\tilde{\gamma}_i) \tag{4.27}$$

[see also Sutradhar and Qu (1998)], where $\tilde{\gamma} = (\tilde{\gamma}_1,\ldots,\tilde{\gamma}_i,\ldots,\tilde{\gamma}_K)$ with $\tilde{\gamma}_i$ as the posterior mode of γ_i computed from

$$\frac{\partial h(\gamma_i)}{\partial \gamma_i} = 0,$$

where

$$h(\gamma_i) = -\sum_{j=1}^{n_i} y_{ij}(x'_{ij}\beta + \gamma_i) + \sum_{j=1}^{n_i} \exp(x'_{ij}\beta + \gamma_i) + \frac{\gamma_i^2}{2\sigma_\gamma^2}.$$

Estimating equations for β and γ_i: Similar to the best linear unbiased prediction (BLUP) approach [Henderson (1950); see also Searle, Casella, and McCulloch (1992, Section 3.4)], the PQL approach pretends that the random effects γ_i are fixed effects parameters and estimate/predict them along with β, before estimating σ_γ^2. The estimating equations for β and γ_i are obtained by maximizing the penalized quasi-likelihood function $[-\sum_{i=1}^{K} h(\gamma_i)]$, with respect to β and γ_i, and they are given by

$$g_1^*(\beta,\gamma_i) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}[y_{ij} - \exp(x'_{ij}\beta + \gamma_i)]x_{ij} = 0 \tag{4.28}$$

and

$$g_2^*(\beta,\gamma_i,\sigma_\gamma^2) = \sum_{j=1}^{n_i}[y_{ij} - \exp(x'_{ij}\beta + \gamma_i)] - \frac{\gamma_i}{\sigma_\gamma^2} = 0 \tag{4.29}$$

for β and γ_i, respectively, where σ_γ^2 is assumed to be known. Let $\hat{\beta}_{PQL}$ and $\hat{\gamma}_{i,PQL}$ $(i = 1,\ldots,K)$ be the solutions.

Estimating equation for σ_γ^2: For the estimation of this variance parameter, a profile quasi-likelihood function is constructed first, by replacing β and $\tilde{\gamma}_i$ in (4.27) with $\hat{\beta}_{PQL}$ and $\hat{\gamma}_{i,PQL}$ $(i = 1,\ldots,K)$, respectively. Next, the profile quasi-likelihood function $ql(\hat{\beta}_{PQL},\sigma_\gamma^2,\hat{\gamma}_{i,PQL})$ is written in the form of a 'working' normal likelihood function and the restricted maximum likelihood estimate of σ_γ^2 is obtained following Patterson and Thompson (1974), for example. This profile quasi-likelihood based score equation for σ_γ^2 is given by

$$\begin{aligned} g_3^*(\hat{\beta}_{PQL},\sigma_\gamma^2,\hat{\gamma}_{i,PQL}) &= \frac{\partial ql(\hat{\beta}_{PQL},\sigma_\gamma^2,\hat{\gamma}_{i,PQL})}{\partial \sigma_\gamma^2} \\ &= \sum_{i=1}^{K}\hat{\gamma}_{i,PQL}^2 - \sigma_\gamma^4 \sum_{i=1}^{K} \frac{\sum_{j=1}^{n_i}\exp(x'_{ij}\hat{\beta}_{PQL} + \hat{\gamma}_{i,PQL})}{1 + \sigma_\gamma^2 \sum_{j=1}^{n_i}\exp(x'_{ij}\hat{\beta}_{PQL} + \hat{\gamma}_{i,PQL})} \\ &= 0. \end{aligned} \tag{4.30}$$

The estimate of σ_γ^2 obtained from (4.30) is denoted by $\hat{\sigma}_{\gamma,PQL}^2$.

Some Remarks on the Asymptotic Properties of the PQL Estimators: Note that it is of interest to estimate only β and σ_γ^2. It is, however, clear from (4.28) and (4.30) that the estimates of these two parameters depend on the estimates of γ_i $(i = 1,\ldots,K)$, where, for a given i, the estimate of γ_i is obtained from (4.29) by exploiting only n_i responses from the ith family. Because n_i is small in the present familial setup, one can only obtain a small sample estimate for this γ_i. Moreover (4.29) also shows that the estimation of γ_i requires the knowledge of σ_γ^2. Consequently, any poor estimates of σ_γ^2 obtained by (4.30) may produce a biased estimate of γ_i for some i and in turn all γ_i $(i = 1,\ldots,K)$ estimates, good or poor, collectively may produce a biased estimate of σ_γ^2. In fact, it may be verified following Sutradhar and Qu (1998) that the normality based profile quasi-likelihood estimating equation (4.30) for σ_γ^2 may not produce a consistent estimate for σ_γ^2, even if one uses true β and true γ_i. The problem will get much worse if a considerable portion of true γ_is are substituted by corresponding biased estimates.

Now to verify the asymptotic property of the estimator of σ_γ^2 obtained from (4.30) for given true values of β and γ_i, we rewrite the equation (4.30) as

$$\sigma_\gamma^2 = \frac{\sum_{i=1}^{K} \gamma_i^2/(K\sigma_\gamma^2)}{(1/K)\sum_{i=1}^{K}[\sum_{j=1}^{n_i} \mu_{ij}^*/(1+\sigma_\gamma^2 \sum_{j=1}^{n_i} \mu_{ij}^*)]}, \tag{4.31}$$

where $\mu_{ij}^* = \exp(x_{ij}'\beta + \gamma_i)$. Note that the true γ_is are iid with zero mean and variance σ_γ^2, thus it follows that

$$\text{limit}_{K\to\infty}(\frac{1}{K})\sum_{i=1}^{K}\gamma_i^2 = \sigma_\gamma^2.$$

Consequently, the right-hand side of (4.31) converges to σ_γ^2, only when $w_i = \sum_{j=1}^{n_i}\mu_{ij}^* = \sum_{j=1}^{n_i}\exp(x_{ij}'\beta + \gamma_i)$ is sufficiently large. For small w_i, the right hand-side of (4.31) converges to a quantity different from σ_γ^2. Thus, the PQL approach may or may not a yield consistent estimate for σ_γ^2, depending on the family size and the covariate information x_{ij} for $j = 1,\ldots,n_i$.

Note that a simulation study reported by Sutradhar and Qu (1998, Table 2, p. 183) also supports the above finding with regard to the poor performance of the PQL approach in estimating σ_γ^2. To be specific, for $K = 100$ families with family effects $\gamma_i^*(i = 1,\ldots,K)$ generated independently from a normal distribution with mean 0 and variance 1, these authors have generated count responses for family members with family size $n_i = 4,6$, by following the Poisson mixed model (4.1) – (4.2) with

$$\eta_{ij}(\gamma_i^*) = \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \beta_4 x_{ij4} + \sigma_\gamma \gamma_i^*,$$

where regression effects were considered to be

$$\beta_1 = 2.5, \quad \beta_2 = -1.0, \quad \beta_3 = 1.0, \quad \text{and} \quad \beta_4 = 0.5,$$

and for all $i = 1,\ldots,K$, covariates were chosen as

$$x_{ij1} = 1.0, \quad \text{for } j = 1,\ldots,n_i,$$

$$x_{ij2} = \begin{cases} 1 & \text{for } j = 1,\ldots,n_i/2 \\ 0 & \text{for } j = (n_i/2)+1,\ldots,n_i, \end{cases}$$

$$x_{ij3} = j - \frac{n_i+1}{2}, \quad \text{for } j = 1, \ldots, n_i, \text{ and}$$

$$x_{ij4} = x_{ij2}x_{ij3}.$$

By using the generated count data $\{y_{ij}\}$ and the above covariates, the regression effects $\beta = [\beta_1,\ \beta_2,\ \beta_3,\ \beta_4]'$ and the variance of the family effects σ_γ^2, were estimated iteratively by solving (4.28) – (4.30). These PQL estimates were obtained for 5000 simulations. The simulated mean (SM) and the standard error (SSE) for the 5000 PQL estimates of σ_γ^2 were found to be as in the following table. It is clear from Table 4.1 that the PQL approach in general overestimates the random effects variance σ_γ^2. The SM value appears to be far away from the true value of σ_γ^2, especially when the true value of σ_γ^2 is large. Thus, the bias appears to get larger as the true value of σ_γ^2 increases. The increase in family size from 4 to 6 helped in bias reduction but the bias still remains very high, showing the inconsistency of the PQL approach for σ_γ^2 estimation, especially when the true value of σ_γ^2 is more than 0.5.

Table 4.1 Simulated means and simulated standard errors of the PQL estimates for σ_γ^2 based on 5000 simulations.

		PQL Estimates for σ_γ^2				
Family Size	Statistic	True $\sigma_\gamma^2 = 0.10$	0.25	0.50	0.75	1.00
4	SM	0.140	0.320	0.649	1.020	1.429
	SSE	0.009	0.016	0.027	0.040	0.057
6	SM	0.119	0.294	0.615	0.975	1.374
	SSE	0.005	0.009	0.015	0.021	0.028

Note that there also exist some bias correction approaches to reduce the biases of the PQL estimators. But these approaches, for example, the bias correction approach discussed in Breslow and Lin (1995, p. 90) for the binary mixed models, appear to improve the results when the true value of σ_γ^2 is small such as less than or equal to 0.50. See also Jiang (1998) for some discussions on the drawbacks of the PQL approach in estimating the variance of the random family effects.

4.2.3 Small Variance Asymptotic Approach: A Likelihood Approximation (LA)

Realizing the difficulties encountered by the PQL approach in consistently estimating the parameters of the Poisson familial models, in particular the variance of the random family effects, some authors such as Sutradhar and Qu (1998) have approximated the Poisson−normal mixed model based likelihood by a Poisson−gamma mixed model based likelihood, the approximation being valid for $\sigma_\gamma^2 \downarrow 0$. It is demonstrated by these authors that this small σ_γ^2 based likelihood approximation performs well in estimating σ_γ^2 even when the true value of σ_γ^2 is as large as 1.0, and it always produces a less biased estimate for σ_γ^2 than the PQL approach.

The following lemma is useful to develop the proposed LA.

Lemma 4.2. Recall from (4.17) that $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma_\gamma^2)$. Let $w_i = \exp(\gamma_i)$. For $\sigma_\gamma^2 \downarrow 0$, the normal density of γ_i can be approximated by a gamma 'working' distribution for w_i given as

$$h_w(w_i) = \frac{\phi^\alpha}{\Gamma(\alpha)} \exp(-\phi w_i) w_i^{\alpha-1}, \tag{4.32}$$

where

$$\alpha = \frac{1}{\exp(\sigma_\gamma^2) - 1} \quad \text{and} \quad \phi = \frac{1}{\exp(\sigma_\gamma^2/2)\{\exp(\sigma_\gamma^2) - 1\}}.$$

Proof: It is easy to prove the lemma in a reverse way. Because $\gamma_i = \log w_i$, the gamma 'working' distribution (4.32) is equivalent to the 'working' distribution of γ_i given by

$$g_w(\gamma_i) = \frac{\phi^\alpha}{\Gamma(\alpha)} \exp\{\alpha\gamma_i - \phi \exp(\gamma_i)\}, \tag{4.33}$$

for any $\sigma_\gamma^2 > 0$. Now, for small σ_γ^2, that is, for $\sigma_\gamma^2 \downarrow 0$, one may use Taylor's series expansion and approximate the probability density of γ_i in (4.33) by a normal density with 0 mean and variance σ_γ^2. $\diamond$

We now use Lemma 4.2 to approximate the exact log likelihood function in (4.18) as follows. First, for $\gamma_i = \sigma_\gamma \gamma_i^* = \log w_i$, we re-express $s_i(\gamma_i^*)$ in (4.20) as

$$s_i^*(w_i) = \left[\log w_i \sum_{j=1}^{n_i} y_{ij} - w_i \sum_{j=1}^{n_i} \exp(x_{ij}'\beta)\right]. \tag{4.34}$$

Next by Lemma 4.2, by using the equivalence of $N(0, \sigma_\gamma^2)$ density for γ_i to the gamma 'working' density $h_w(w_i)$ in (4.32), we write an approximation to the integral in (4.19) as

$$\begin{aligned} J_i &\approx \int_{-\infty}^{\infty} \exp[s_i^*(w_i)] h_w(w_i) dw_i \\ &= J_i^* \text{ (say)}. \end{aligned} \tag{4.35}$$

Note that this integral is computable and it yields a log likelihood approximation to (4.18) as

$$\begin{aligned} \log L(\beta, \sigma_\gamma) \approx &- \sum_{i=1}^{K} \sum_{j=1}^{n_i} \log y_{ij}! + \sum_{i=1}^{K} \sum_{j=1}^{n_i} y_{ij} x_{ij}'\beta + \sum_{i=1}^{K} \log \frac{\Gamma(\alpha + \sum_{j=1}^{n_i} y_{ij})}{\Gamma(\alpha)} \\ &- \sum_{i=1}^{K} \left(\alpha + \sum_{j=1}^{n_i} y_{ij}\right) \log \left(\phi + \sum_{j=1}^{n_i} \exp(x_{ij}'\beta)\right) + K\alpha \log \phi \\ = &\log L^*(\beta, \sigma_\gamma) \text{ (say)}. \end{aligned} \tag{4.36}$$

One may then solve the following likelihood estimating equations for β and σ_γ^2 given by

$$U_1^*(\beta,\sigma_\gamma^2) = \frac{\partial \log L^*(\beta,\sigma_\gamma)}{\partial\beta} = \sum_{i=1}^K \left(\sum_{j=1}^{n_i} y_{ij}x_{ij} - \frac{y_i^*}{\mu_i^*}\sum_{j=1}^{n_i} x_{ij}\exp(x'_{ij}\beta) \right) = 0, \quad (4.37)$$

and

$$U_2^*(\beta,\sigma_\gamma^2) = \frac{\partial \log L^*(\beta,\sigma_\gamma)}{\partial\sigma_\gamma^2} = \alpha'(\sigma_\gamma^2)\sum_{i=1}^K \left(\psi(y_i^*) - \psi(\alpha) + \log\frac{\phi}{\mu_i^*} \right) + \phi'(\sigma_\gamma^2)\sum_{i=1}^K \left(\frac{\alpha}{\phi} - \frac{y_i^*}{\mu_i^*} \right) = 0, \quad (4.38)$$

respectively, where

$$y_i^* = \alpha + \sum_{j=1}^{n_i} y_{ij}, \quad \mu_i^* = \phi + \sum_{j=1}^{n_i}\exp(x'_{ij}\beta), \quad \alpha'(\sigma_\gamma^2) = \frac{\partial\alpha}{\partial\sigma_\gamma^2} = -\frac{\exp(\sigma_\gamma^2)}{[\exp(\sigma_\gamma^2)-1]^2},$$

and

$$\phi'(\sigma_\gamma^2) = \frac{\partial\phi}{\partial\sigma_\gamma^2} = -\frac{3\exp(\sigma_\gamma^2)-1}{2\exp(\sigma_\gamma^2/2)[\exp(\sigma_\gamma^2)-1]^2}.$$

Let $\hat{\beta}_{LA}$ and $\hat{\sigma}^2_{\gamma,LA}$ be the LA estimates obtained from (4.37) – (4.38) for β and σ_γ^2, respectively.

To examine the relative performances of the LA and PQL estimation approaches for β and σ_γ^2, Sutradhar and Qu (1998) have used the same simulated data that we have described in the last section and obtained 500 values of $\hat{\beta}_{LA}$ and $\hat{\sigma}^2_{\gamma,LA}$. The average and standard errors of these 500 estimates are available in Tables 4.1 and 4.2 in Sutradhar and Qu (1998). For convenience, we show the simulation results for the estimator $\hat{\sigma}^2_{\gamma,LA}$ in Table 4.2 below. The results of this table under the LA approach correspond to the results of Table 4.1 under the PQL approach discussed in Section 4.2.

Table 4.2 Simulated means and simulated standard errors of the LA estimates for σ_γ^2 based on 5000 simulations.

		LA Estimates for σ_γ^2				
Family Size	Statistic	True $\sigma_\gamma^2 = 0.10$	0.25	0.50	0.75	1.00
4	SM	0.102	0.244	0.436	0.591	0.722
	SSE	0.007	0.009	0.010	0.011	0.011
6	SM	0.101	0.242	0.434	0.588	0.718
	SSE	0.004	0.005	0.006	0.007	0.007

When the results of Table 4.2 are compared to those of Table 4.1, it is clear that the LA estimates (Table 4.2) of σ_γ^2 are much closer to the true values of σ_γ^2 than the corresponding PQL estimates (Table 4.1). For example, when the family size is $n_i = 4$, the LA estimate for true $\sigma_\gamma^2 = 0.5$ is 0.436, whereas the PQL estimate was 0.649, showing that the PQL approach is more biased than the LA approach. The simulated standard errors appear to be much more stable and smaller under the LA approach as compared to the PQL approach.

As far as the performances of the PQL and LA approaches for the estimation of β are concerned, the LA approach performs much better than the PQL approach in estimating the intercept parameter β_1, whereas they perform almost the same for the estimation of other regression parameters, namely, β_2, β_3, and β_4. See Table 4.3 below, for example, for a simulation based comparative performance of the PQL and LA approaches when σ_γ^2 is small. Their relative performances in estimating the regression effects, for other small and moderately large values of σ_γ^2, can be found in Table 1 in Sutradhar and Qu (1998, p. 181).

Table 4.3 Simulated means and simulated standard errors of the LA and PQL estimates for regression effects when $\sigma_\gamma^2 = 0.25$ based on 5000 simulations.

			Estimates for Regression Effects			
Family Size	Method	Statistic	$\beta_1 = 2.5$	$\beta_2 = -1.0$	$\beta_3 = 1.0$	$\beta_4 = 0.5$
4	LA	SM	2.422	−0.993	1.000	0.519
		SSE	0.033	0.126	0.024	0.156
	PQL	SM	2.283	−0.997	1.00	0.510
		SSE	0.026	0.126	0.024	0.156
6	LA	SM	2.425	−0.994	1.000	0.518
		SSE	0.022	0.103	0.010	0.112
	PQL	SM	2.207	−0.999	1.000	0.509
		SSE	0.017	0.103	0.010	0.112

Remark that when the combined results of Tables 4.1 and 4.3 are examined, the PQL approach appears to overestimate σ_γ^2 (Table 4.1) and underestimate β_1 (Table 4.3). Thus, it appears that this approach is able to properly estimate $\beta_1 + \frac{1}{2}\sigma_\gamma^2$ as a confounded effect involved in the exponent for the unconditional mean $\mu_{ij} = \exp(x'_{ij}\beta + \frac{1}{2}\sigma_\gamma^2)$ (4.14) of the response y_{ij}, whereas the LA approach is able to produce almost unbiased estimates for β_1 and σ_γ^2 separately and hence does not suffer from any identification of the parameter problems.

4.2.3.1 A Higher-Order Likelihood Approximation (HOLA)

Recall that in the LA approach, for $\sigma_\gamma^2 \downarrow 0$, the true normal density of γ_i, that is,

$$g_N(\gamma_i|\sigma_\gamma^2) = (2\pi\sigma_\gamma^2)^{-1/2}\exp\{-\gamma_i{}^2/2\sigma_\gamma^2\}$$

was approximated by $g_w(\gamma_i)$, a 'working' density of γ_i as given by (4.33). The main objective of the HOLA approach is to approximate the true density $g_N(\gamma_i)$ of γ_i

by a better density than $g_w(\gamma_i)$, in order to obtain a better likelihood function than $L^*(\beta,\sigma_\gamma)$ in (4.36). Let $\tilde{g}_w(\gamma_i)$ denote the improved probability density which will have its first four moments the same as the first four moments of the true distribution $g_N(\gamma_i)$, whereas $g_w(\gamma_i)$ has its first two moments the same as the first two moments of the true distribution $g_N(\gamma_i)$. The improved likelihood function based on the improved density $\tilde{g}_w(\gamma_i)$ of γ_i is denoted by $\tilde{L}(\beta,\sigma_\gamma)$.

In order to derive the improved density $\tilde{g}_w(\gamma_i)$ as a function of $g_w(\gamma_i)$ (4.33), we use the well-known Gram−Charlier series expansion [cf. Johnson and Kotz (1972, pp. 15 − 22)] up to the term with moments up to order four, and obtain

$$\tilde{g}_w(\gamma_i) \approx \Big[1-\varepsilon_1 P_1(\gamma_i)+\frac{1}{2}\{\varepsilon_1^2+\varepsilon_2\}P_2(\gamma_i)-\frac{1}{6}\{\varepsilon_1^2+3\varepsilon_1\varepsilon_2+\varepsilon_3\}P_3(\gamma_i) + \frac{1}{24}\{\varepsilon_1^4+6\varepsilon_1^2\varepsilon_2+4\varepsilon_1\varepsilon_3+\varepsilon_4\}P_4(\gamma_i)\Big]\, g_w(\gamma_i), \tag{4.39}$$

where, for $\ell = 1,\ldots,4$,

$$\varepsilon_\ell = K_\ell - K_\ell^*, \tag{4.40}$$

with K_ℓ and K_ℓ^* as the ℓth cumulants of the distributions $g_N(\gamma_i)$ and $g_w(\gamma_i)$, respectively, and the formulas for $P_\ell(\gamma_i)$ are obtained by writing the ℓth derivative of $g_w(\gamma_i)$ with respect to γ_i^*, in the form

$$\frac{\partial^\ell g_w(\gamma_i)}{\partial {\gamma_i}^\ell} = P_\ell(\gamma_i) g_w(\gamma_i). \tag{4.41}$$

See Exercise 4.1 for the specific values of K_ℓ and K_ℓ^* and Exercise 4.2 for the formulas for $P_\ell(\gamma_i)$, for all $\ell = 1,\ldots,4$.

Note that the formula for $g_w(\gamma_i)$ in (4.39) is given by (4.33) which is approximated by $h_w(w_i) = [\phi^\alpha/\Gamma(\alpha)]\exp(-\phi w_i)w_i^{\alpha-1}$ as given in (4.32), where $w_i = \exp(\gamma_i)$. Now express $\tilde{g}_w(\gamma_i)$ in (4.39) as

$$\tilde{g}_w(\gamma_i) = \tilde{h}_w(w_i) = P^*(w_i)h_w(w_i), \tag{4.42}$$

where $P^*(w_i)$ is a function of w_i, and is equivalent to the function in the square bracket in the right-hand side of (4.39), and $h_w(w_i)$ is written for $g_w(\gamma_i)$. Consequently, the integral J_i in (4.19) is now approximated as

$$\begin{aligned} J_i &\approx \int_{-\infty}^{\infty} \exp[s_i^*(w_i)]\tilde{h}_w(w_i)dw_i \\ &= \tilde{J}_i \text{ (say)}, \end{aligned} \tag{4.43}$$

where $s_i^*(w_i)$ is given by (4.34). Next by using this $\tilde{J}_i$ for J_i in (4.18), one obtains an improved log likelihood approximation over (4.36), given by

$$\log \tilde{L}(\beta,\sigma_\gamma) = -\sum_{i=1}^{K}\sum_{j=1}^{n_i}\log y_{ij}! + \sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}x_{ij}'\beta + K\alpha\log\phi$$

$$+\sum_{i=1}^{K} log\left[\sum_{r=1}^{5}\frac{C_r\Gamma(y^*_{ir})}{\Gamma(\alpha)(\mu^*_i)^{y^*_{ir}}}\right], \tag{4.44}$$

[Sutradhar and Das (2001, eqn. (3.3), p. 63] where, for

$$\alpha=\frac{1}{\exp(\sigma_\gamma^2)-1} \quad \text{and} \quad \phi=\frac{1}{\exp(\sigma_\gamma^2/2)\{\exp(\sigma_\gamma^2)-1\}},$$

$$y^*_{ir}=\alpha+r-1+\sum_{j=1}^{n_i}y_{ij} \text{ and } \mu^*_i=\phi+\sum_{j=1}^{n_i}\exp(x'_{ij}\beta),$$

and

$$\begin{aligned}
C_1 &= 1-\frac{1}{6}\alpha^3\sigma_\gamma^4-\frac{1}{12}\alpha^4\sigma_\gamma^6\\
C_2 &= \frac{1}{6}\phi[1+3\alpha+3\alpha^2]\sigma_\gamma^4+\frac{1}{12}\phi[1+4\alpha+6\alpha^2+4\alpha^3]\sigma_\gamma^6\\
C_3 &= -\frac{1}{2}\phi^2[1+\alpha]\sigma_\gamma^4-\frac{1}{12}\phi^2[7+12\alpha+6\alpha^2]\sigma_\gamma^6\\
C_4 &= \frac{1}{6}\phi^3\sigma_\gamma^4+\frac{1}{6}\phi^3[3+2\alpha]\sigma_\gamma^6\\
C_5 &= -\frac{1}{12}\phi^4\sigma_\gamma^6.
\end{aligned}$$

One then uses (4.44) to write the likelihood estimating equations for β and σ_γ^2 given by

$$\tilde{U}_1 = \frac{\partial\log\tilde{L}(\beta,\sigma_\gamma)}{\partial\beta}=\sum_{i=1}^{K}\left[\sum_{j=1}^{n_i}y_{ij}x_{ij}-\frac{s_{i2}}{s_{i1}}\sum_{j=1}^{n_i}x_{ij}\exp(x'_{ij}\beta)\right]=0 \tag{4.45}$$

$$\begin{aligned}
\tilde{U}_2 &= \frac{\partial\log\tilde{L}(\beta,\sigma_\gamma)}{\partial\sigma_\gamma^2}=\left[K\alpha'(\sigma_\gamma^2)\{\log\phi-\psi(\alpha)\}+K\alpha\phi'(\sigma_\gamma^2)/\phi\right]\\
&\quad+\sum_{i=1}^{K}\frac{s'_{i1}(\sigma_\gamma^2)}{s_{i1}}=0,
\end{aligned} \tag{4.46}$$

where $\psi(\alpha)=\partial\log\Gamma(\alpha)/\partial\alpha$, and

$$s_{i1}=\sum_{r=1}^{5}C_rp_{ir},\ \ s_{i2}=\sum_{r=1}^{5}C_rp_{i,r+1},$$

$$s'_{i1}(\sigma_\gamma^2)=\sum_{r=1}^{5}\left[C'_r(\sigma_\gamma^2)p_{ir}+C_rp_{ir}w_{ir}\right],$$

with C_r as in (4.44) and $C'_r(\sigma_\gamma^2)=\partial C_r/\partial\sigma_\gamma^2$, and

$$p_{ir} = \Pi_{u=1}^{r-1} \frac{y_{iu}^*}{\{\mu_i^*\}^{r-1}}$$

$$w_{ir} = \left[\alpha'(\sigma_\gamma^2)\{\psi(y_{ir}^*) - \log \mu_i^*\} - \phi'(\sigma_\gamma^2)\frac{y_{ir}^*}{\mu_i^*}\right].$$

Let $\hat{\beta}_{HOLA}$ and $\hat{\sigma}^2_{\gamma,HOLA}$ be the improved higher-order likelihood approximate estimates for β and σ_γ^2 obtained from (4.45) and (4.46), respectively.

Note that it is expected that the HOLA approach will yield better estimates than the LA approach. In order to have a quantitative idea on the relative performances of these two approaches, Sutradhar and Das (2001, Table 4.1) conducted a simulation study using $K = 100$ families each with $n_i = 6$ members. For simplicity, as opposed to 4 covariates considered by Sutradhar and Qu (1998), Sutradhar and Das (2001) have considered $p = 2$ covariates. These two covariates were chosen as

$$x_{ij1} = \begin{cases} 1 & \text{for } j = 1,\ldots,n_i/2; \quad i = 1,\ldots,K/2, \\ 0 & \text{for } j = n_i/2+1,\ldots,n_i; \; i = 1,\ldots,K, \\ 1 & \text{for } j = 1,\ldots,n_i; \quad i = K/2+1,\ldots,K \end{cases}$$

$$x_{ij2} = \begin{cases} 1 & \text{for } j = 1,\ldots,n_i/2; \quad i = 1,\ldots,K/2, \\ 2 & \text{for } j = n_i/2+1,\ldots,n_i; \quad i = 1,\ldots,K, \\ -1 & \text{for } j = 1,\ldots,n_i/3; \quad i = K/2+1,\ldots,K, \\ 0 & \text{for } j = n_i/3+1,\ldots,2n_i/3; \; i = K/2+1,\ldots,K, \\ 1 & \text{for } j = 2n_i/3+1,\ldots,n_i; \quad i = K/2+1,\ldots,K. \end{cases}$$

For selected true values of β_1, β_2, and σ_γ^2, by generating data as in Section 4.2, the LA estimates of these parameters were obtained by solving (4.37) and (4.38), and their HOLA estimates were obtained from (4.45) and (4.46), respectively. The simulated means and standard errors of the estimates based on 1000 simulations are shown in Table 4.4 below.

The results of Table 4.4 show that in estimating both β_1 and β_2, in general, the HOLA approach leads to a considerable bias reduction as compared to the LA approach. For example, the true $\beta_1 = 1.0$ was estimated by the LA approach as 0.953, 0.951, and 0.953 when $\sigma_\gamma^2 = 0.60,\ 0.75,\ 0.90$, respectively, whereas the corresponding HOLA estimates for β_1 were found to be 0.984, 0.993, and 0.983. Similarly, the HOLA approach leads to a significant improvement in estimating σ_γ^2 as compared to the LA approach. For example, the true $\sigma_\gamma^2 = 0.60$ was estimated as 0.487 and 0.537 by the LA and HOLA approaches, respectively. The HOLA estimator appears to perform very well when true σ_γ^2 is large. Thus, in general, the higher-order

Table 4.4 Simulated means and simulated standard errors of the LA and HOLA estimates for regression effects as well as σ_γ^2, based on family size $n_i = 6$, for $K = 100$ families, and 1000 simulations, when $\beta_1 = \beta_2 = 1.0$.

			True σ_γ^2			
Method	Estimates	Statistic	0.40	0.60	0.75	0.90
LA	$\hat{\beta}_{LA,1}$	SM	0.949	0.953	0.951	0.953
		SSE	0.029	0.025	0.021	0.019
	$\hat{\beta}_{LA,2}$	SM	0.948	0.946	0.947	0.946
		SSE	0.018	0.016	0.014	0.013
	$\hat{\sigma}^2_{\gamma,LA}$	SM	0.385	0.487	0.675	0.637
		SSE	0.015	0.015	0.014	0.014
HOLA	$\hat{\beta}_{HOLA,1}$	SM	0.925	0.984	0.993	0.983
		SSE	0.086	0.023	0.019	0.018
	$\hat{\beta}_{HOLA,2}$	SM	0.955	0.957	0.961	0.951
		SSE	0.053	0.015	0.013	0.013
	$\hat{\sigma}^2_{\gamma,HOLA}$	SM	0.417	0.537	0.675	0.811
		SSE	0.008	0.017	0.006	0.003

likelihood approximation leads to significant improvement over the LA approach in estimating all β and σ_γ^2 parameters of the model.

4.2.4 Hierarchical Likelihood (HL) Approach

Similarly to the PQL approach, there exists a hierarchical likelihood (HL) approach [Lee and Nelder (1996)] that uses the estimates of $\gamma_i (i = 1, \ldots, K)$ to estimate the desired regression parameter β and the overdispersion parameter σ_γ^2. The difference between the two approaches is that the PQL approach estimates β and γ_i by solving their estimating equations (4.28) and (4.29) developed by maximizing a penalized quasi-likelihood function, whereas the HL approach maximizes the hierarchical likelihood function

$$\mathbf{h} = \log \prod_{i=1}^{K} \prod_{j=1}^{n_i} f(y_{ij}|\gamma_i, \beta) + \log \prod_{i=1}^{K} g(\gamma_i|\sigma_\gamma^2), \tag{4.47}$$

to estimate these parameters, where $g(\gamma_i|\sigma_\gamma^2)$ is the density function of unobserved γ_i, and $f(y_{ij}|\gamma_i, \beta)$ is the Poisson density function as in (4.1) for the response y_{ij} given γ_i. Similarly to the PQL approach, we use the normal density

$$g_N(\gamma_i|\sigma_\gamma^2) = (2\pi\sigma_\gamma^2)^{-1/2} \exp\{-\gamma_i{}^2/2\sigma_\gamma^2\}$$

for $g(\gamma_i|\sigma_\gamma^2)$ in (4.47). Note that as far as the estimation of σ_γ^2 is concerned, the PQL approach solves the profile quasi-likelihood function based estimating equation (4.30), whereas the HL approach maximizes an adjusted profile hierarchical likelihood function [Lee and Nelder (1996)] given by

$$h_A = \mathbf{h} + \frac{1}{2}\log\{\det(2\pi H^{-1})\}$$
$$= \mathbf{h} + \frac{1}{2}\log(2\pi) - \frac{1}{2}\log\{\det(H)\}, \quad (4.48)$$

where $\mathbf{h}$ is the hierarchical likelihood function as in (4.47) and the H matrix is defined as

$$H = \begin{bmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ+U \end{bmatrix}_{(p+K)\times(p+K)}, \quad (4.49)$$

where $X = [X_1, ..., X_i, ..., X_K]' : \sum n_i \times p$, with $X_i = [x_{i1}, \ldots, x_{ij}, \ldots, x_{in_i}]'$ as the $n_i \times p$ covariate matrix; W and Z are block-diagonal matrices given by $W = \bigoplus_{i=1}^K A_i^* : \sum n_i \times \sum n_i$ and $Z = \bigoplus_{i=1}^K 1_{n_i} : \sum n_i \times K$, respectively, with $A_i^* = \text{diag}[\mu_{i1}^*, ..., \mu_{in_i}^*]$: $\sum n_i \times \sum n_i$ where $\mu_{ij}^* = \exp(x'_{ij}\beta + \gamma_i)$; and $1_{n_i} = (1, ..., 1)' : \sum n_i \times 1$; and $U = [1/\sigma_\gamma^2]I_K$, I_K being the $K \times K$ identity matrix.

Note that by maximizing the HL function in (4.47) with respect to β and γ_i, one obtains the HL estimating functions for β and γ_i given by

$$\frac{\partial \mathbf{h}}{\partial \beta} = \sum_{i=1}^K X_i'(y_i - \mu_i^*) = 0, \quad (4.50)$$
$$\frac{\partial \mathbf{h}}{\partial \gamma_i} = \sum_{j=1}^{n_i}(y_{ij} - \mu_{ij}^*) - \frac{\gamma_i}{\sigma_\gamma^2} = 0. \quad (4.51)$$

Next, for the HL estimation of σ_γ^2, the maximization of h_A, the adjusted profile HL function in (4.48), is achieved by using the iterative equation given by

$$\hat{\sigma}^2_{\gamma(r+1)} = \hat{\sigma}^2_{\gamma(r)} + \left[\left(\frac{\partial^2 h_A}{\partial \sigma_\gamma^4}\right)^{-1} \frac{\partial h_A}{\partial \sigma_\gamma^2}\right]_{(r)}, \quad (4.52)$$

where the square bracket $[\,]_{(r)}$ indicates that the quantity in [ ] is evaluated at $\sigma_\gamma^2 = \hat{\sigma}^2_{\gamma(r)}$, r being the rth iteration. In (4.52),

$$\frac{\partial h_A}{\partial \sigma_\gamma^2} = -\frac{K}{2\sigma^2} + \frac{\sum_{i=1}^K \gamma_i^2}{2\sigma_\gamma^4} + \frac{tr(D)}{2\sigma_\gamma^4}, \quad (4.53)$$
$$\frac{\partial^2 h_A}{\partial \sigma_\gamma^4} = \frac{K}{2\sigma_\gamma^4} - \frac{\sum_{i=1}^K \gamma_i^2}{\sigma_\gamma^6} - \frac{tr(D)}{\sigma_\gamma^6} + \frac{tr(DD)}{2\sigma_\gamma^8}, \quad (4.54)$$

with D = $[(Z'WZ+U) - Z'WXX'WXX'WZ]^{-1}$ as the bottom diagonal matrix of H^{-1} with appropriate dimension, H being defined in (4.49).

Let $\hat{\beta}_{HL}$ be the solution of the HL based estimating equation (4.50) for β, and $\hat{\sigma}^2_{\gamma,HL}$ be obtained as the HL based estimate of σ_γ^2 from the iterative equation (4.52). A simulation study was conducted by Chowdhury and Sutradhar (2009) in order to examine the relative performances of these HL estimators to the corresponding generalized quasi-likelihood (GQL) estimators suggested by Sutradhar (2004). In Section 4.6, we provide details on this GQL estimation approach and discuss its superior performance over the HL approach based on a simulation study by Chowdhury and Sutradhar (2009).

4.2.5 Method of Moments (MM)

As opposed to the best linear unbiased prediction analogue such as the PQL (Breslow and Clayton, 1993) and the HL (Lee and Nelder, 1996) approaches, Jiang (1998) discussed a simulated method of moments (SMM) for the estimation of the parameters of the generalized linear mixed models, the binary and Poisson mixed models being two important special cases. The main purpose of the introduction of a simpler method of moments is to handle the multidimensional random effects cases fairly easily as compared to other approaches. In this approach, the normal random effects are driven out by simulating them first and then numerically averaging over their distributions. These are done to compute the necessary unconditional moments of the responses, whereas the PQL and HL approaches estimate the parameters of the models through the prediction/estimation of the random effects. The moment approach is also expected to produce consistent estimates for the parameters, whereas the PQL and HL approaches may fail to produce such consistent estimates especially for the variance component of the random effects.

Note that as opposed to the binary case, one does not need any simulation or other numerical techniques to compute the first− or higher-order moments of the count responses following Poisson mixed models. Thus, we provide here the estimating equations based on an ordinary method of moments (MM) where necessary moments can be computed directly provided random effects are normally distributed. More specifically, under the Poisson mixed models, one estimates β and σ^2_γ by solving the moment equations

$$\psi_1(\beta,\sigma^2_\gamma)=\sum_{i=1}^{K}\sum_{j=1}^{n_i}x_{ij}\left\{y_{ij}-\mu_{ij}(\beta,\sigma^2_\gamma)\right\}=0, \tag{4.55}$$

and

$$\psi_2(\beta,\sigma^2_\gamma)=\sum_{i=1}^{K}\left[\left(\sum_{j=1}^{n_i}y_{ij}\right)^2-\left(\sum_{j=1}^{n_i}\lambda_{ijj}(\beta,\sigma^2_\gamma)+2\sum_{j<k}^{n_i}\lambda_{ijk}(\beta,\sigma^2_\gamma)\right)\right]=0, \tag{4.56}$$

respectively, where

$$\mu_{ij}=E[Y_{ij}]=\exp(x'_{ij}\beta+\frac{1}{2}\sigma^2_\gamma),\quad \lambda_{ijj}=E[Y^2_{ij}]=\mu_{ij}+\exp(\sigma^2_\gamma)\mu^2_{ij},$$

and

$$\lambda_{ijk} = E[Y_{ij}Y_{ik}] = \exp(\sigma_\gamma^2)\mu_{ij}\mu_{ik},$$

by $(4.14) - (4.16)$.

Now by re-expressing these two moment equations in (4.55) and (4.56) as

$$\begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} = \begin{bmatrix} w_1 - \xi_1 \\ w_2 - \xi_2 \end{bmatrix} = 0, \tag{4.57}$$

and writing

$$\psi = [\psi_1', \psi_2]', \quad w = [w_1', w_2]', \text{ and } \xi = [\xi_1', \xi_2]',$$

one may obtain the MM estimate of $\theta = [\beta', \sigma_\gamma^2]'$ by using the Gauss$-$Newton iterative equation

$$\hat{\theta}_{MM}(r+1) = \hat{\theta}_{MM}(r) + \left[\frac{\partial \xi'}{\partial \theta}\right]_r^{-1} [w - \xi]_r, \tag{4.58}$$

where $[]_r$ denotes that the expression within the square bracket is evaluated at $\theta = \hat{\theta}_{MM}(r)$, the estimate obtained for the rth iteration. Let the final solution obtained from (4.58) be denoted by $\hat{\theta}_{MM}$.

Note that because $E[\psi] = 0$, the MM estimator $\hat{\theta}_{MM}$ is consistent for θ but it may still produce biased estimators in finite sample cases. Moreover, the MM estimator can be inefficient. In the next section, it is demonstrated through a simulation study that the MM approach indeed produces highly biased estimates in the finite sample cases for both regression and variance component parameters, especially when the true variance parameter value is large. The simulation study also includes a moments based generalized quasi-likelihood approach proposed by Sutradhar (2004) which appears to work much better than the MM approach, in estimating all regression and variance component parameters.

4.2.6 Generalized Quasi-Likelihood (GQL) Approach

Let $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$ be the n_i response vector collected from n_i members of the ith $(i = 1, \ldots, K)$ family. Next, write the mean vector of y_i and its covariance matrix as

$$E[Y_i] = \mu_i(\beta, \sigma_\gamma^2) = (\mu_{i1}, \ldots, \mu_{ij}, \ldots, \mu_{in_i})' : n_i \times 1 \tag{4.59}$$

$$\text{Cov}[Y_i] = \Sigma_i(\beta, \sigma_\gamma^2) = (\sigma_{jk}) : n_i \times n_i, \tag{4.60}$$

where, by Lemma 4.1,

$$\begin{aligned} \mu_{ij} &= \exp(x_{ij}'\beta + \frac{1}{2}\sigma_\gamma^2) \\ \sigma_{ijj} &= \text{var}[Y_{ij}] = \mu_{ij} + [\exp(\sigma_\gamma^2) - 1]\mu_{ij}^2 \end{aligned}$$

$$\sigma_{ijk} = \text{cov}[Y_{ij}, Y_{ik}] = \mu_{ij}\mu_{ik}[\exp(\sigma_\gamma^2) - 1], \text{ for } j \neq k.$$

To understand the nature of the data through this Poisson mixed model, it is of interest to estimate the parameters β and σ_γ^2. Note that $\sigma_\gamma^2 = 0$ would reduce the mixed model (4.1) – (4.2) to a fixed model. Also it implies that the responses of the members would be independent. Now, if the responses were independent, one could have estimated the only parameter β by using the well-known quasi-likelihood (QL) approach of Wedderburn (1974) [see also McCullagh (1983)] which exploits the means and variances of the data. More specifically, the QL estimating equation would have been

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[\frac{\partial \mu'_{ij}}{\partial \beta}\frac{(y_{ij}-\mu_{ij})}{\text{var}(y_{ij})}\right] = 0, \tag{4.61}$$

with $\mu_{ij} = \exp(x'_{ij}\beta)$.

4.2.6.1 Marginal Generalized Quasi-Likelihood (GQL) Estimation of β

For the correlated responses, Sutradhar (2003, Section 3) has proposed a generalization of the QL approach of Wedderburn (1974) to a longitudinal setup, where the mean vector and covariance matrix of the responses are utilized in estimating the parameter(s) involved in the mean vector. Furthermore, Sutradhar (2004) has used this generalized quasi-likelihood approach in the binary and Poisson familial setup. For the present Poisson familial (mixed) model with the unconditional mean vector μ_i and covariance matrix Σ_i as in (4.59) and (4.60), respectively, the GQL estimating equation for β, assuming known σ_γ^2, is given by

$$\sum_{i=1}^{K}\frac{\partial \mu'_i}{\partial \beta}\Sigma_i^{-1}(y_i - \mu_i) = 0. \tag{4.62}$$

The expectation of the estimating function in the left-hand side of (4.62) is zero, therefore the GQL estimator, say $\hat{\beta}_{GQL}$ obtained by solving (4.62) would be consistent for β. Furthermore, because the estimating equation (4.62) is fully standardized by using the inverse of the covariance matrix as the weight matrix, $\hat{\beta}_{GQL}$ will also be highly efficient, the maximum likelihood estimator being fully efficient or optimal, which is, however, not easy to obtain under the present mixed model setup.

Note that the solution of (4.62) may be obtained by using the Gauss–Newton iterative equation

$$\hat{\beta}_{GQL}(r+1) = \hat{\beta}_{GQL}(r) + \left[\sum_{i=1}^{K}\frac{\partial \mu'_i}{\partial \beta}\Sigma_i^{-1}\frac{\partial \mu_i}{\partial \beta'}\right]_r^{-1}\left[\sum_{i=1}^{K}\frac{\partial \mu'_i}{\partial \beta}\Sigma_i^{-1}(y_i - \mu_i)\right]_r, \tag{4.63}$$

where $[]_r$ denotes that the expression within the square bracket is evaluated at $\beta = \hat{\beta}_{GQL}(r)$, the estimate obtained for the rth iteration. It also can be shown that asymptotically (as $K \to \infty$), for known σ_γ^2, the final GQL estimator obtained from (4.63) follows the multivariate Gaussian distribution with mean β and the covariance matrix given by

$$\mathrm{cov}(\hat{\beta}_{GQL}) = \mathrm{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1} \frac{\partial \mu_i}{\partial \beta'}\right]^{-1}. \tag{4.64}$$

4.2.6.2 Marginal Generalized Quasi-Likelihood (GQL) Estimation of σ_γ^2

Sutradhar [2004, eqn. (3.4), p. 270] has developed the GQL estimating equations for the joint estimation of β and σ_γ^2. It can be shown that, conditional on γ_i^*, all first− and second-order responses can form a sufficient statistic for the parameters β and σ_γ^2 involved in the generalized linear function $\eta_{ij} = x_{ij}'\beta + \sigma_\gamma \gamma_i^*$ (Jiang, 1998), therefore we write a second-order response based GQL estimating equation for the marginal estimation of σ_γ^2, whereas the first-order responses were used to construct the marginal GQL estimating equation (4.62) for the estimation of β.

Let

$$u_i = (u_{i1}', u_{i2}')' \tag{4.65}$$

be the vector of all second-order responses under the ith family, where

$$\begin{aligned} u_{i1} &= (y_{i1}^2, \ldots, y_{ij}^2, \ldots, y_{in_i}^2)' : n_i \times 1, \\ u_{i2} &= (y_{i1}y_{i2}, \ldots, y_{ij}y_{ik}, \ldots, y_{i(n_i-1)}y_{in_i})', \;\; j<k : \frac{n_i(n_i-1)}{2} \times 1. \end{aligned}$$

Furthermore, let

$$\begin{aligned} \lambda_i &= E[U_i] \\ &= (\lambda_{i11}, \ldots, \lambda_{ijj}, \ldots, \lambda_{in_in_i}, \lambda_{i12}, \ldots, \lambda_{ijk}, \ldots, \lambda_{i(n_i-1)n_i})', \end{aligned} \tag{4.66}$$

where, by (4.56), $\lambda_{ijj} = E[Y_{ij}^2] = \mu_{ij} + \exp(\sigma_\gamma^2)\mu_{ij}^2$ for all $j = 1, \ldots, n_i$, and $\lambda_{ijk} = E[Y_{ij}Y_{ik}] = \exp(\sigma_\gamma^2)\mu_{ij}\mu_{ik}$ for all $j \neq k, j,k = 1, \ldots, n_i$. Also, let

$$\Omega_i = \mathrm{cov}(U_i) = \begin{bmatrix} \mathrm{cov}(U_{i1}) & \mathrm{cov}(U_{i1}, U_{i2}') \\ & \mathrm{cov}(U_{i2}) \end{bmatrix} \tag{4.67}$$

$$= \begin{bmatrix} F_i & G_i \\ & H_i \end{bmatrix}, \tag{4.68}$$

where the formulas for the component matrices F_i, G_i, and H_i, are given in Lemma 4.3 below. In the fashion similar to that of (4.62), for known β, one may solve the

GQL estimating equation

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma_\gamma^2} \Omega_i^{-1} (u_i - \lambda_i) = 0, \tag{4.69}$$

for σ_γ^2 in order to obtain a consistent and highly efficient estimator for σ_γ^2. Let $\hat{\sigma}^2_{\gamma,GQL}$ denote the solution of (4.69) which can be obtained by using the iterative equation

$$\begin{aligned} \hat{\sigma}^2_{\gamma,GQL}(r+1) &= \hat{\sigma}^2_{\gamma,GQL}(r) + \left[\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma_\gamma^2} \Omega_i^{-1} \frac{\partial \lambda_i}{\partial \sigma_\gamma^2} \right]_r^{-1} \\ &\quad \times \left[\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma_\gamma^2} \Omega_i^{-1} (u_i - \lambda_i) \right]_r, \end{aligned} \tag{4.70}$$

where $[]_r$ denotes that the expression within the square bracket is evaluated at $\sigma_\gamma^2 = \hat{\sigma}^2_{\gamma,GQL}(r)$, the estimate obtained for the rth iteration. Furthermore, similar to that of (4.64), it can be shown that asymptotically (as $K \to \infty$), for known β, the final GQL estimator obtained from (4.70) follows the univariate Gaussian distribution with mean σ_γ^2 and the variance given by

$$\text{var}(\hat{\sigma}^2_{\gamma,GQL}) = \text{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma_\gamma^2} \Omega_i^{-1} \frac{\partial \lambda_i}{\partial \sigma_\gamma^2} \right]^{-1}. \tag{4.71}$$

Note that in practice, the iterative equations (4.63) for β and (4.70) for σ_γ^2 constitute a cycle, and the cycles of operation continues until convergence, to obtain the final GQL estimates $\hat{\beta}_{GQL}$ and $\hat{\sigma}^2_{\gamma,GQL}$ for β and σ_γ^2, respectively.

Lemma 4.3. Recall from (4.1) that conditional on the random family effect γ_i, y_{ij} follows the Poisson distribution with mean parameter $\mu_{ij}^* = \exp(x_{ij}'\beta + \gamma_i)$, for all $j = 1, \ldots, n_i$. Also we have assumed that $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma_\gamma^2)$. For $u_i = (u_{i1}', u_{i2}')'$ as in (4.65), it then follows that the formulas for the component matrices of $\Omega_i = \text{cov}[U_i]$, namely of F_i, G_i, and H_i (4.68), are given by

Formula for cov$[U_{i1}] = F_i$

$$\begin{aligned} \text{var}[Y_{ij}^2] &= \phi_{ijjjj} - \lambda_{ijj}^2 \\ &= \mu_{ij} \left[1 + 7\mu_{ij} \exp(\sigma_\gamma^2) + 6\mu_{ij}^2 \exp(3\sigma_\gamma^2) \right. \\ &\quad \left. + \mu_{ij}^3 \exp(6\sigma_\gamma^2) \right] - \lambda_{ijj}^2, \text{ for } j = 1, \ldots, n_i \end{aligned} \tag{4.72}$$

$$\begin{aligned} \text{cov}[Y_{ij}^2, Y_{ik}^2] &= \phi_{ijjkk} - \lambda_{ijj}\lambda_{ikk} \\ &= \mu_{ij}\mu_{ik} \exp(\sigma_\gamma^2) \left[1 + \{\mu_{ij} + \mu_{ik}\} \exp(2\sigma_\gamma^2) \right. \end{aligned}$$

$$+\,\mu_{ij}\mu_{ik}\exp(5\sigma_\gamma^2)\big] - \lambda_{ijj}\lambda_{ikk},\ \text{for } j \neq k,\ j,k = 1,\ldots,n_i, \tag{4.73}$$

where the formulas for μ_{ij} and λ_{ijj} for all $j = 1,\ldots,n_i$, are given by (4.56).

Formula for cov$[U_{i2}] = H_i$

$$\text{var}[Y_{ij}Y_{ik}] = \phi_{ijjkk} - \lambda_{ijk}^2 \ \text{ for } j \neq k \tag{4.74}$$

$$\text{cov}[Y_{ij}Y_{ik}, Y_{i\ell}Y_{im}] = \begin{cases} \phi_{ijjkm} - \lambda_{ijk}\lambda_{ijm} & \text{for } j = \ell \\ \phi_{ijjk\ell} - \lambda_{ijk}\lambda_{ij\ell} & \text{for } j = m \\ \phi_{ijkkm} - \lambda_{ijk}\lambda_{ikm} & \text{for } k = \ell \\ \phi_{ijkk\ell} - \lambda_{ijk}\lambda_{ik\ell} & \text{for } k = m \end{cases} \tag{4.75}$$

$$\text{cov}[Y_{ij}Y_{ik}, Y_{i\ell}Y_{im}] = \phi_{ijk\ell m} - \lambda_{ijk}\lambda_{i\ell m} \tag{4.76}$$

$$= \mu_{ij}\mu_{ik}\mu_{i\ell}\mu_{im}\exp(6\sigma_\gamma^2) - \lambda_{ijk}\lambda_{i\ell m},\ \text{ for } j \neq \ell,\ k \neq m, \tag{4.77}$$

where λ_{ijk} for $j \neq k$ are given by (4.56), ϕ_{ijjkk} is given in (4.73), and

$$\phi_{ijjkm} = \mu_{ij}\mu_{ik}\mu_{im}\exp(3\sigma_\gamma^2)\left[1 + \mu_{ij}\exp(3\sigma_\gamma^2)\right]$$

$$\phi_{ijkkm} = \mu_{ij}\mu_{ik}\mu_{im}\exp(3\sigma_\gamma^2)\left[1 + \mu_{im}\exp(3\sigma_\gamma^2)\right],$$

for example.

Formula for cov$[U_{i1}, U'_{i2}] = G_i$

$$\text{cov}[Y_{ij}^2, Y_{ik}Y_{i\ell}] = \begin{cases} \phi_{ijjj\ell} - \lambda_{ijj}\lambda_{ij\ell} & \text{for } j = k \\ \phi_{ijjjk} - \lambda_{ijj}\lambda_{ijk} & \text{for } j = \ell \end{cases} \tag{4.78}$$

$$\text{cov}[Y_{ij}^2, Y_{ik}Y_{i\ell}] = \phi_{ijjk\ell} - \lambda_{ijj}\lambda_{ik\ell},\ \text{ for } j \neq k, j \neq \ell, \tag{4.79}$$

where $\phi_{ijjk\ell}$ is as in (4.75), and

$$\phi_{ijjjk} = \mu_{ij}\mu_{ik}\exp(\sigma_\gamma^2)[1 + 3\mu_{ij}\exp(2\sigma_\gamma^2) + \mu_{ij}^2\exp(5\sigma_\gamma^2)],$$

for example.

Proof: All these formulas for the moments can be derived by computing first the appropriate conditional moments for given random family effect γ_i, and then taking the average over the distribution of the random family effects. For example, to compute $\phi_{ijjjj} = E(Y_{ij}^4)$, by (4.11),we first compute the corresponding condition moment as

$$E(Y_{ij}^4|\gamma_i) = \mu_{ij}^* + 7\mu_{ij}^{*2} + 6\mu_{ij}^{*3} + \mu_{ij}^{*4},$$

where $\mu_{ij}^* = \exp(x_{ij}'\beta + \gamma_i)$. We then take the expectation of this conditional moment over the normal distribution for γ_i, and obtain

$$\phi_{ijjjj} = E[Y_{ij}^4] = E_{\gamma_i} E[Y_{ij}^4]$$
$$= \mu_{ij}\left[1 + 7\mu_{ij}\exp(\sigma_\gamma^2) + 6\mu_{ij}^2\exp(3\sigma_\gamma^2) + \mu_{ij}^3\exp(6\sigma_\gamma^2)\right]$$

as in (4.72). Similarly, the remaining fourth-order moments are computed using the basic steps as follows;

$$\begin{aligned}
\phi_{ijjkk} &= E_{\gamma_i}[E(Y_{ij}^2|\gamma_i)E(Y_{ik}^2|\gamma_i)] \\
&= E_{\gamma_i}[\{\mu_{ij}^* + \mu_{ij}^{*2}\}\{\mu_{ik}^* + \mu_{ik}^{*2}\}] \\
\phi_{ijjjk} &= E_{\gamma_i}[E(Y_{ij}^3|\gamma_i)E(Y_{ik}|\gamma_i)] \\
&= E_{\gamma_i}[\{\mu_{ij}^* + 3\mu_{ij}^{*2} + \mu_{ij}^{*3}\}\{\mu_{ik}^*\}] \\
\phi_{ijjk\ell} &= E_{\gamma_i}[E(Y_{ij}^2|\gamma_i)E(Y_{ik}|\gamma_i)E(Y_{i\ell}|\gamma_i)] \\
&= E_{\gamma_i}[\{\mu_{ij}^* + \mu_{ij}^{*2}\}\mu_{ik}^*\mu_{i\ell}^*] \\
\phi_{ijk\ell} &= E_{\gamma_i}[E(Y_{ij}|\gamma_i)E(Y_{ik}|\gamma_i)E(Y_{i\ell}|\gamma_i)E(Y_{im}|\gamma_i)] \\
&= E_{\gamma_i}[\mu_{ij}^*\mu_{ik}^*\mu_{i\ell}^*\mu_{im}^*]. \quad \diamond
\end{aligned}$$

4.2.6.3 Joint Generalized Quasi-Likelihood (GQL) Estimation for β and σ_γ^2

For quick convergence of the estimates, one may like to estimate β and σ_γ^2 jointly. For this, the estimating equations (4.62) and (4.69) may be combined as follows. Let

$$s_i = (y_i', u_i')'$$

with

$$E[S_i] = \zeta_i = (\mu_i', \lambda_i')', \text{ and } \text{cov}[S_i] = \Upsilon_i, \tag{4.80}$$

where μ_i and λ_i are as in (4.62) and (4.66), respectively, and

$$\Upsilon_i = \text{cov}(S_i) = \begin{bmatrix} \text{cov}(Y_i) & \text{cov}(Y_i, U_i') \\ & \text{cov}(U_i) \end{bmatrix} \tag{4.81}$$

$$= \begin{bmatrix} \Sigma_i & \Lambda_i \\ & \Omega_i \end{bmatrix}, \tag{4.82}$$

with Σ_i and Ω_i as in (4.62) and (4.68),respectively, and

$$\Lambda_i = [\text{cov}(Y_i, U'_{i1}) \;\; \text{cov}(Y_i, U'_{i2})] = [B_i \;\; E_i].$$

The formulas for B_i and E_i may be computed as follows.

Formula for cov$[Y_i, U'_{i1}] = B_i$

$$\begin{aligned} \text{cov}[Y_{ij}, Y_{ik}^2] &= \delta_{ijkk} - \mu_{ij}\lambda_{ijk} \\ &= E_{\gamma_i}[E(Y_{ij}|\gamma_i)E(Y_{ik}^2|\gamma_i)] - \mu_{ij}\lambda_{ijk} \\ &= E_{\gamma_i}[\{\mu^*_{ij}\}\{\mu^*_{ik} + \mu^{*2}_{ik}\}] - \mu_{ij}\lambda_{ijk} \qquad (4.83) \\ &= \begin{cases} \mu_{ij}\left[1 + 3\mu_{ij}\exp(\sigma^2_\gamma) + \mu^2_{ij}\exp(3\sigma^2_\gamma) \right. \\ \left. -\mu_{ij}\{1 + 3\mu_{ij}\exp(\sigma^2_\gamma)\}\right] & \text{for } j = k \\ \mu_{ij}\mu_{ik}\left[\{\exp(\sigma^2_\gamma) - 1\} \right. \\ \left. +\mu_{ik}\exp(\sigma^2_\gamma)\{\exp(2\sigma^2_\gamma) - 1\}\right] & \text{for } j \neq k. \end{cases} \end{aligned}$$

Formula for cov$[Y_i, U'_{i2}] = E_i$

$$\begin{aligned} \text{cov}[Y_{ij}, Y_{ik}Y_{i\ell}] &= \delta_{ijk\ell} - \mu_{ij}\lambda_{ik\ell}, \; k \neq \ell \\ &= E_{\gamma_i}[E(Y_{ij}|\gamma_i)E(Y_{ik}|\gamma_i)E(Y_{i\ell}|\gamma_i)] - \mu_{ij}\lambda_{ik\ell} \\ &= E_{\gamma_i}[\mu^*_{ij}\mu^*_{ik}\mu^*_{i\ell}] - \mu_{ij}\lambda_{ik\ell} \qquad (4.84) \\ &= \begin{cases} \mu_{ij}\mu_{i\ell}\exp(\sigma^2_\gamma)\left[1 + \mu_{ij}\{\exp(2\sigma^2_\gamma) - 1\}\right] & \text{for } j = k \\ \mu_{ij}\mu_{ik}\exp(\sigma^2_\gamma)\left[1 + \mu_{ij}\{\exp(2\sigma^2_\gamma) - 1\}\right] & \text{for } j = \ell \\ \mu_{ij}\mu_{ik}\mu_{i\ell}\exp(\sigma^2_\gamma)\left[\exp(2\sigma^2_\gamma) - 1\right] & \text{for } j \neq k, \; j \neq \ell. \end{cases} \end{aligned}$$

For $\theta = (\beta', \sigma^2_\gamma)'$, it then follows that the joint GQL estimating equation for β and σ^2_γ may be written as

$$\sum_{i=1}^{K} \frac{\partial \zeta'_i}{\partial \theta} \Upsilon_i^{-1}(s_i - \zeta_i) = 0. \tag{4.85}$$

This equation can be solved by using the iterative equation

$$\hat{\theta}_{GQL}(r+1) = \hat{\theta}_{GQL}(r) + \left[\sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \theta} \Upsilon_i^{-1} \frac{\partial \zeta_i}{\partial \theta'}\right]_r^{-1} \left[\sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \theta} \Upsilon_i^{-1} (s_i - \zeta_i)\right]_r, \quad (4.86)$$

where $[]_r$ denotes that the expression within the square bracket is evaluated at $\theta = \hat{\theta}_{GQL}(r)$, the estimate obtained for the rth iteration. Furthermore, similar to that of (4.71), it can be shown that asymptotically (as $K \to \infty$), the final GQL estimator obtained from (4.86) follows the multivariate Gaussian distribution with mean θ and the variance given by

$$\text{var}(\hat{\theta}_{GQL}) = \text{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \theta} \Upsilon_i^{-1} \frac{\partial \zeta_i}{\partial \theta'}\right]^{-1}. \quad (4.87)$$

4.2.7 Efficiency Comparison

4.2.7.1 Efficiency Comparison Between GQL and MM Approaches: A Small Sample Study

We now examine the efficiency performance of the GQL and MM estimators through a simulation study. For simplicity, we consider a Poisson mixed model with two fixed covariates and one source of random effects, so that conditional on the random effects, the count response is generated from the Poisson distribution with mean

$$\mu_{ij}^* = \exp(x_{ij1}\beta_1 + x_{ij2}\beta_2 + \sigma_\gamma \gamma_i^*), \quad (4.88)$$

[see (4.3)]. We consider $\beta_1 = \beta_2 = 1$ and $K = 100$ clusters. Furthermore, we consider $n_i = n = 6$ for all i. The two design covariates were chosen as

$$x_{ij1} = \begin{cases} 1 & \text{for } j = 1,\ldots,3;\ i = 1,\ldots,K/2 \\ 0 & \text{for } j = 4,\ldots,6;\ i = 1,\ldots,K/2 \\ 1 & \text{for } j = 1,\ldots,6;\ i = (K/2)+1,\ldots,K; \end{cases}$$

and

$$x_{ij2} = \begin{cases} 1 & \text{for } j = 1,\ldots,3;\ i = 1,\ldots,K/2 \\ 2 & \text{for } j = 4,\ldots,6;\ i = 1,\ldots,K/2 \\ -1 & \text{for } j = 1,2;\ i = (K/2)+1,\ldots,K \\ 0 & \text{for } j = 3,4;\ i = (K/2)+1,\ldots,K \\ 1 & \text{for } j = 5,6;\ i = (K/2)+1,\ldots,K. \end{cases}$$

In addition, the γ_i were independently generated from a standard normal distribution. With regard to the selection of the variance of the random effects, we choose σ_γ^2 = 0.4, 0.8, 1.0, and 1.25. We remark that even though in theory the overdispersion index parameter σ_γ^2 can take any value from 0 to ∞, for practical purposes $\sigma_\gamma^2 \geq 1.0$ appears to be quite large. This is because under the Poisson-normal mixed model (4.1), the overdispersion in the count data may increase significantly even if the increment in σ_γ^2 is small. To be specific, the variance of y_{ij}, $\sigma_{ijj} = \mu_{ij} + [\exp(\sigma_\gamma^2) - 1]\,\mu_{ij}^2$, under the Poisson-normal mixed model increases significantly, depending on the value of the mean function $\mu_{ij} = \exp(x'_{ij}\beta + \frac{1}{2}\sigma_\gamma^2)$, even if σ_γ^2 changes from 1.0 to 1.2, for example. We further remark that Breslow and Lin (1995, P. 90) were able to obtain unbiased estimates of this overdispersion index parameter σ_γ^2 when σ_γ^2 ranges only up to 0.5.

To simulate the data, the responses $(y_{i1}, \ldots, y_{in_i})$ for $n_i = 6$ for each cluster i were generated as realizations of the Poisson model (4.1) with mean and variance equal to $\mu_{ij}^* = \exp(\beta_1 x_{ij1} + \beta_2 x_{ij2} + \sigma_\gamma \gamma_i^*)$. The simulated data (y_{ij}), $j = 1, \ldots, 6$, $i = 1, \ldots, K$ $(K = 100)$, and the covariates (x_{iju}), $u = 1, \ldots, p$; $j = 1, \ldots, 6$, $i = 1, \ldots, K$, were used to compute the estimates of the fixed-effect parameters β and variance component σ_γ^2 of the random effects, based on the MM and GQL approaches discussed in Sections 4.2.5 and 4.2.6.3, respectively. More specifically, the estimates were obtained by using the Newton−Raphson iterative equation (4.58) to solve (4.57) for the moment estimates and by using (4.86) to solve (4.85) for the joint GQL estimates. We have used the same small initial values for each of the β and σ_γ^2 parameters under both moment and GQL approaches. The iterative procedure was terminated when the difference between the estimates of two consecutive iterations was less than or equal to 0.005. The simulation was repeated 1000 times in order to obtain the mean value and standard errors of the parameter estimates.

Table 4.5 Comparison of the MM and GQL based simulated mean values, standard errors, and mean squared errors of the regression estimates for selected values of σ_γ^2; $K = 100$; $n_i = n = 6$ $\beta_1 = \beta_2 = 1$; 1000 simulations.

Regression parameter	Method	Quantity	Regression Estimates $\sigma_\gamma^2 = 0.40$	0.80	1.00	1.25
β_1	MM	Mean	1.048	1.067	1.072	1.079
		SE	0.024	0.026	0.026	0.026
		MSE	0.003	0.005	0.006	0.007
	GQL	Mean	1.016	1.023	1.000	1.035
		SE	0.021	0.029	0.045	0.033
		MSE	0.001	0.001	0.002	0.002
β_2	MM	Mean	0.891	0.815	0.780	0.738
		SE	0.016	0.017	0.017	0.017
		MSE	0.012	0.035	0.049	0.069
	GQL	Mean	0.951	0.952	0.953	0.957
		SE	0.016	0.021	0.022	0.025
		MSE	0.003	0.003	0.003	0.003

Note that β and σ_γ^2 were estimated jointly based on moment and GQL approaches discussed in Sections 4.2.5 and 4.2.6.3. Table 4.5 reports the simulated mean values and standard errors of the estimates of β_1 and β_2 computed by: (1) Jiang's moment method, and (2) the joint generalized quasi-likelihood approach. As both the moment and GQL approaches yield biased estimates, we compute the mean squared errors to study the efficiency of the estimators. More specifically, the efficiency of one estimation method as compared to its counterpart should be computed by comparing the mean squared errors of the estimator, not by comparing the variances of the estimators produced by the methods. With this in view, we report the simulated mean values, standard errors, and the mean squared errors of the regression estimators in Table 4.5 and for the estimator of the variance component in Table 4.6.

Table 4.6 Comparison of the MM and GQL based simulated mean values, standard errors, and mean squared errors of the estimates of variance components of the random effects for selected values of σ_γ^2; $K = 100$; $n_i = n = 6$; $\beta_1 = \beta_2 = 1$; 1000 simulations.

		Variance Component Estimate			
Method	Quantity	$\sigma_\gamma^2 = 0.40$	0.80	1.00	1.25
MM	Mean	0.192	0.410	0.529	0.677
	SE	0.019	0.027	0.032	0.036
	MSE	0.044	0.153	0.223	0.330
GQL	Mean	0.353	0.789	0.990	1.376
	SE	0.042	0.117	0.222	0.225
	MSE	0.004	0.014	0.049	0.067

It is clear from Table 4.5 that in estimating both β_1 and β_2, in general, the GQL approach leads to a large reduction in bias and hence in mean squared errors relative to the moment approach. In particular, the moment approach performs very poorly in estimating β_2 as compared to the GQL approach. For example, for the case when $\sigma_\gamma^2 = 0.8$, the moment approach yields 0.035 as the mean squared error of the estimator of β_2, whereas the GQL approach yields only 0.003 for this estimator, resulting in very large mean squared error efficiency gain for the GQL approach. Note that the simulations are done also for large σ_γ^2 such as $\sigma_\gamma^2 = 1.0, 1.25$, which are beyond the ranges for σ_γ^2 considered by Breslow and Lin (1995) and Sutradhar and Qu (1998). In all cases, the GQL approach performs better than the moment approach.

It is clear from Table 4.6 that for all σ_γ^2, the GQL method performs extremely well in estimating σ_γ^2 as compared to the MM approach of Jiang (1998). The moment approach grossly underestimates σ_γ^2, whether σ_γ^2 is large or small, whereas the GQL approach slightly underestimates σ_γ^2 when σ_γ^2 is small, and overestimates σ_γ^2 when σ_γ^2 is large. But as compared to the moment approach, the amount of bias is relatively insignificant. Furthermore it is clear from the table that the MSEs yielded by the moment approach are much larger than those of the GQL approach. The performance of the moment approach is worse for the large σ_γ^2 cases. This is because as σ_γ^2 increases, the moment approach appears to be highly biased in estimating σ_γ^2, as compared to the cases with smaller σ_γ^2. In general, the standard errors of both GQL and moment estimators increase as σ_γ^2 increases, but the moment approach produces smaller standard errors for the large σ_γ^2 cases. This better performance of the moment approach in producing smaller SE is apparently due to the fact that the moment approach appears to yield similar simulated estimates but they are far off from the actual parameter values. Thus, in summary, the GQL approach performs much better than the moment approach in estimating all parameters of the Poisson-mixed model including the regression effects.

4.2.7.2 Efficiency Comparison Between GQL and HL Approaches: A Small Sample Study

As pointed out in Section 4.2.4, the hierarchical likelihood approach [Lee and Nelder (1996)] is conceptually quite similar to the penalized quasi-likelihood (PQL) approach [Breslow and Clayton (1993)]. Both of these approaches use the predicted random effects for the estimation of the regression effects β and the random effects variance component σ_γ^2. As discussed in Section 4.2.2, it is, however, known that the PQL approach may not produce consistent estimates for σ_γ^2, especially when the true value of σ_γ^2 is large. Because of the similarity between the PQL and HL approaches, the HL approach may also produce biased and hence inconsistent estimates. A simulation study by Chowdhury and Sutradhar (2009) appears to support this conjecture that the HL approach similar to the PQL approach may encounter convergence difficulties in estimating the parameters, especially the variance component of the mixed model. We present here a part of this simulation study by Chowdhury and Sutradhar (2009).

The data were generated in the same way as in the last section with Poisson mean given by (4.88). For the family size, we now consider two values, namely, $n_i = n = 4,6$, for all $i = 1,\ldots,K = 100$. For the variance component of the random effects, we choose $\sigma_\gamma^2 = 0.4, 0.8$, and 1.20. As far as the covariates are concerned, the first covariate is kept the same as in the last section but a slightly different second covariate was chosen. These covariates are:

$$x_{ij1} = \begin{cases} 1 \text{ for } j = 1,2,\ldots,n_i/2; \quad i = 1,2,\ldots,K/2 \\ 0 \text{ for } j = n_i/2+1,\ldots,n_i; \quad i = 1,2,\ldots,K/2 \\ 1 \text{ for } j = 1,\ldots,n_i; \quad i = K/2+1,\ldots,K \end{cases}$$

$$x_{ij2} = \begin{cases} 1 & \text{for } j = 1,2,\ldots,n_i/2; \quad i = 1,2,\ldots,K/2 \\ 2 & \text{for } j = n_i/2+1,\ldots,n_i; \quad i = 1,2,\ldots,K/2 \\ 0 & \text{for } j = 1,2,\ldots,n_i/2; \quad i = K/2+1,\ldots,K \\ 1 & \text{for } j = n_i/2+1,\ldots,n_i; \quad i = K/2+1,\ldots,K \end{cases}$$

Next, under each simulation, the simulated values of $\{y_{ij}\}$ along with the values of the covariates $\{x_{ij}\}$ were used to obtain the HL estimates of β and σ_γ^2 by using (4.50) and (4.52), respectively. To obtain the GQL estimates, unlike in the last section, we have solved the marginal GQL estimating equations (4.62) for β and (4.69) for σ_γ^2. Note that under the HL approach, we also had to estimate γ_i $(i = 1,\ldots,100)$ by treating them as the fixed parameters, but these estimates were not reported as they are not of direct interest. The simulated means (Mean) and simulated standard errors (SE) for the GQL and HL based regression estimates are shown in Table 4.7 for selected cluster sizes $n_i = 4$, and 6, and for all selected values of σ_γ^2.

Note that in the last section, we have used the simulated MSEs for comparing the efficiency of the GQL and MM estimates. This type of MSE based comparison is appropriate when competitive approaches are not so biased but they produce different standard errors. However, when an estimate becomes highly biased with small standard error, it turns out to be an useless estimate. For this reason, to compare the performances of the actual convergence of the estimates to their corresponding parameter values, in this section we have computed the simulated relative bias (RB) given by

$$\text{RB} = \frac{|\text{Mean} - \text{True parameter value}|}{\text{SE}} \times 100.$$

These RBs for the regression estimates are reported in the same Table 4.7 for two cluster sizes and selected values of the overdispersion index parameter.

With regard to the estimation of β_1 and β_2, the results in Table 4.7 show that the GQL approach always produces the regression estimates with smaller relative bias as compared to the HL approach. This better performance of the GQL approach appears to hold for both cluster sizes n_i = 4, and 6; as well as for all small and large values of σ_γ^2 = 0.4, 0.8, and 1.2. For example, when n_i = 4 and σ_γ^2 = 0.4, the GQL estimates of β_1 and β_2 are slightly biased with RBs 27 and 40, and respective RBs are 15 and 22 when σ_γ^2 = 1.2. But, the HL estimates for the same regression parameters appear to converge to wrong values with small standard errors. To be specific, for n_i = 4, the HL estimates of β_1 and β_2 appear to have RBs 300 and 411 when σ_γ^2 = 0.4, and strikingly large RBs 1162 and 1582 when σ_γ^2 = 1.2. For cluster size $n_i = 6$, the performance of the HL based regression estimation appears to improve for large variance components, but the RBs still remain higher than for the corresponding GQL estimates. Thus, irrespective of the cluster size n_i and the value of σ_γ^2, the GQL approach performs much better than the HL approach in estimating β_1 and β_2.

Note that the comparison between GQL and HL approaches by Chowdhury and Sutradhar (2009) was done for wide-ranging values for family size, with smallest

Table 4.7 Comparison of the GQL and HL based simulated mean values, standard errors, and relative biases of the regression estimates for selected values of σ_γ^2; $K = 100$; $\beta_1 = \beta_2 = 1$; 500 simulations.

Family Size (n_i)	Regression Parameter	Method	Quantity	Regression Estimates $\sigma_\gamma^2 = 0.40$	0.80	1.20
4	β_1	HL	Mean	1.0878	1.1661	1.2894
			SE	0.0293	0.0277	0.0249
			RB	300	600	1162
		GQL	Mean	1.0109	1.0125	1.0079
			SE	0.0405	0.0483	0.0523
			RB	27	26	15
	β_2	HL	Mean	1.0760	1.1430	1.2594
			SE	0.0185	0.0205	0.0164
			RB	411	698	1582
		GQL	Mean	1.0122	1.0108	1.0097
			SE	0.0305	0.0399	0.0442
			RB	40	27	22
6	β_1	HL	Mean	1.0958	1.1583	0.6222
			SE	0.0240	0.0270	0.1131
			RB	399	586	334
		GQL	Mean	1.0112	1.0081	1.0049
			SE	0.0351	0.0460	0.0491
			RB	32	18	10
	β_2	HL	Mean	1.0831	1.1366	0.6498
			SE	0.0152	0.0207	0.1039
			RB	547	660	337
		GQL	Mean	1.0115	1.0080	1.0053
			SE	0.0283	0.0375	0.0415
			RB	41	21	13

size $n_i = 2$ and the largest size $n_i = 16$. It was found by these authors that the pattern of the regression estimates as a function of n_i and σ_γ^2 is different under the HL approach as compared to the GQL approach. When cluster size is small such as n_i = 2, 4, and 6, the RBs of the regression estimates were found to get smaller for the GQL estimates but they were found to get larger for the HL estimates, as the value of σ_γ^2 increases. When cluster size is large such as n_i = 10 and 16, the performances of the GQL estimates of the regression parameters were found to remain the same irrespective of the value of σ_γ^2, whereas the HL estimates were found to perform better as the value of σ_γ^2 increases. But, when HL and GQL approaches are compared, the GQL approach was found to perform uniformly better than the HL approach in estimating β_1 and β_2.

With regard to the estimation of the overdispersion parameter σ_γ^2, the results in Table 4.8 show that the GQL approach, in general, performs better than the HL approach. For example, when n_i = 6, the GQL approach produces σ_γ^2 estimates with RBs 14 and 23 for σ_γ^2 = 0.4 and 1.2, respectively; whereas the corresponding RBs for the HL estimates are found to be 65 and 196, respectively.

Table 4.8 Comparison of the HL and GQL based simulated mean values, standard errors, and mean squared errors of the estimates of variance components of the random effects for selected values of σ_γ^2; $K = 100$; $\beta_1 = \beta_2 = 1$; 500 simulations.

			Variance Component Estimate		
Family Size (n_i)	Method	Quantity	$\sigma_\gamma^2 = 0.40$	0.80	1.20
4	HL	Mean	0.4043	0.8281	1.3488
		SE	0.0547	0.0528	0.0727
		RB	8	53	205
	GQL	Mean	0.3923	0.7819	1.1719
		SE	0.0646	0.0773	0.0898
		RB	12	23	31
6	HL	Mean	0.4170	0.8499	1.8340
		SE	0.0261	0.0467	0.3234
		RB	65	107	196
	GQL	Mean	0.3926	0.7939	1.1806
		SE	0.0539	0.1449	0.0847
		RB	14	4	23

4.2.8 A Health Care Data Utilization Example

As an application of the familial count data model we consider a dataset on health care utilization, collected by the Department of Community Medicine, Health Science Center (General Hospital) St. John's, Canada. This dataset consists of information on the number of visits paid to a physician during 1985 by 180 members of 48 families. Also information on various associated covariates such as gender, education level, chronic disease condition, and age were collected. This familial data-set is a part of the complete familial longitudinal data collected from the members of these 48 families over a period of six years from 1985 to 1990. The complete dataset is given in Table 6A in the appendix of Chapter 6. However, our purpose here is to study the familial data for a given year, such as 1985. Note that in the present set up the responses are counts. Furthermore, as n_i (three or four) members belong to the same family, it is likely that the responses of the family members would be correlated. These correlations are referred to as the structural correlations. It is of scientific interest to take the structural correlations into account and examine the effects of selected covariates on the number of visits paid by a member to the physician.

We consider four important associated covariates: gender (x_{ij1}), the chronic condition (x_{ij2})[CC], education level (x_{ij3})[EL], and age of the individual (x_{ij4}); and code them as follows.

$$x_{ij1} = \begin{cases} 0 & \text{female} \\ 1 & \text{male} \end{cases} \qquad x_{ij2} = \begin{cases} 0 & \text{without chronic diseases} \\ 1 & \text{with chronic diseases} \end{cases}$$

$$x_{ij3} = \begin{cases} 0 & \text{less than high school} \\ 1 & \text{high school or above} \end{cases} \qquad x_{ij4} = \text{ exact age of the individual.}$$

However, before we consider a formal analysis for the effects of the covariates on the count responses, it is helpful to understand the summary statistics for the data. For this purpose, we present the observed distribution of the count responses, from 180 individuals, by all four covariates in Table 4.9.

Table 4.9 Summary statistics of physician visits by four covariates in the Health Care Utilization Data for 1985.

		Number of Visits					
Covariates	**Level**	**0**	**1**	**2**	$3-5$	$\geq$**6**	**Total**
Gender	Male	28	22	18	16	12	96
	Female	11	5	15	21	32	84
Chronic Condition	No	26	20	15	16	11	88
	Yes	13	7	18	21	33	92
Education Level	$<$ High School	17	5	11	10	15	58
	$\geq$ High School	22	22	22	27	29	122
Age	$20-30$	23	17	14	15	15	84
	$31-40$	1	1	3	3	3	11
	$41-50$	4	4	5	12	8	33
	$51-65$	10	5	8	5	13	41
	$66-85$	1	0	3	2	5	11

It is seen from Table 4.9 that, in general, more males appear to visit their physician a smaller number of times, whereas a large number of females visit the physician at least three times. As expected, we see that an individual with chronic diseases visits a physician more often. Physician visits for individuals with a higher level of education seems to be evenly distributed, that is, individuals are just as likely to visit a physician once as three to five times. For those with a lower level of education, they appear to either not visit their physician, or visit a large number of times. With regard to the relationship between number of visits and age, we have temporarily made five age groups and observed that some of the individuals in the $20-30$ age group have visited a physician a large number of times. As expected, a large number of individuals did not visit a physician at all. For older age groups, there was a tendency for an individual to see the physician more often.

We now turn back to the confirmatory analysis. The main objective is to find the effects of the aforementioned four covariates on the physician visits by the members of 48 randomly chosen families. Jowaheer, Sutradhar, and Sneddon (2009, Section 4.1.2), among other things, computed these effects by using the MM and joint GQL approaches. Chowdhury and Sutradhar (2009) reanalyzed this dataset by using the HL and marginal GQL approaches. The GQL estimates reported in these two works were found to provide similar results, except that because of different coding for some covariates such as gender, the numerical values for the estimates were different. For convenience, we now provide here the MM, HL, and marginal GQL estimates from Chowdhury and Sutradhar (2009) but interpret the GQL estimates only as it is evident by the simulation studies discussed in Sections 4.2.7.1 and 4.2.7.2 that the GQL approach produces less biased and more efficient estimates than the

MM and HL approaches. Note that the random effects variance is an index parameter for familial/structural correlations among the count responses of the members of a given family. As discussed in the last section, this variance component or familial correlation index parameter plays an important role in obtaining consistent and efficient estimates for the effects of the covariates. The estimates for regression effects and variance component along with their estimated standard errors are displayed in Table 4.10.

Table 4.10 The MM, HL, and marginal GQL estimates along with the corresponding estimated standard errors, for the Health Care Utilization Data for 1985.

		Effects of the Covariates				Variance
Method	**Quantity**	**Gender($\hat{\beta}_1$)**	**CC($\hat{\beta}_2$)**	**EL($\hat{\beta}_3$)**	**Age($\hat{\beta}_4$)**	$\hat{\sigma}^2_\gamma$
Marginal GQL	**Value**	−0.754	0.666	0.434	0.010	0.873
	SE	0.091	0.125	0.123	0.0030	0.409
HL	**Value**	−0.693	0.689	0.633	0.016	0.187
	SE	0.080	0.088	0.067	0.0017	0.020
MM	**Value**	−0.651	0.686	0.511	0.014	0.529
	SE	0.079	0.088	0.067	0.0017	—

First, the large value of the GQL estimate of σ^2_γ, that is, $\hat{\sigma}^2_\gamma = 0.873$, indicates that the data is highly overdispersed. This parameter estimate appears to explain the basic mean and variance of the observed data very well. This is because, when we computed the mean and the variance of the count responses from all 180 members, it was found that on the average each individual member visited a physician 3.92 times with very large variance 22.66. Further note that the variance component also affects the familial correlations.

Next, with regard to the regression effects, the negative value of $\hat{\beta}_{1(GQL)}$, namely $\hat{\beta}_{1(GQL)} = -0.754$ indicates that females made more visits to the physician as compared to males. The positive values for $\hat{\beta}_{2(GQL)}$ and $\hat{\beta}_{4(GQL)}$, namely, $\hat{\beta}_{2(GQL)} = 0.666$ and $\hat{\beta}_{4(GQL)} = 0.010$ suggest that the individuals having some chronic diseases or individuals who are older pay more visits to the physician, as expected. The effect of the education level on the health condition, however, appears to be intriguing. This is because $\hat{\beta}_{3(GQL)} = 0.434$ suggests that highly educated individuals have more visits compared to individuals with a lower level of education. One of the possible reasons for this type of behavior of this covariate may be that individuals with a higher level of education are more concerned about their health condition and also they have better facilities as compared to the individuals with a lower level of education.

4.3 Estimation for Multiple Random Effects Based Parametric Mixed Models

4.3.1 Random Effects in a Two-Way Factorial Design Setup

Consider the Poisson mixed model in (4.1) but suppose that unlike (4.2), η_{ij} is now a function of more than one random effect. To be specific, let γ_i be the same random effect as in (4.2) such that $\gamma_i \overset{\text{iid}}{\sim} (0, \sigma_\gamma^2)$. Furthermore, let $\alpha_{ij} \overset{\text{iid}}{\sim} (0, \sigma_\alpha^2)$ denote the individual random effect of the jth member of the ith family. Then for $\gamma_i^* = \gamma_i/\sigma_\gamma$ and $\alpha_{ij}^* = \alpha_{ij}/\sigma_\alpha$, the conditional mean in a two-way design setup for the count responses following the model (4.1), may be expressed as

$$E[Y_{ij}|\gamma_i, \alpha_{ij}] = \mu_{ij}^* = \exp[\eta_{ij}], \tag{4.89}$$

[Lin (1997), Jiang (1998), and Sutradhar and Rao (2003)] with

$$\eta_{ij} = h(x_{ij}'\beta + \sigma_\gamma \gamma_i^* + \sigma_\alpha \alpha_{ij}^*), \tag{4.90}$$

$h(\cdot)$ being a known link function, and β the effect of the covariates x_{ij} on the response y_{ij}. It is clear from the model (4.1) with η_{ij} in (4.90) that the responses $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ under the ith family are influenced by both an unobservable random family effect as well as by an unobservable individual random effects.

As far as the inferences for the parameters β, σ_γ^2, and σ_α^2 are concerned, Sutradhar and Rao (2003) have demonstrated in the context of familial binary data analysis that the generalized quasi-likelihood approach produces consistent and highly efficient estimates as compared to the so-called method of moments considered by Jiang (1998). In a manner similar to that of Sutradhar and Rao (2003), one may deal with the Poisson mixed model defined by (4.1) and (4.90), and develop the GQL and MM estimating equations for all three parameters.

4.3.2 One-Way Heteroscedastic Random Effects

As opposed to the two-way random effects models defined by (4.1) and (4.90), there also exist one way random effects models with heteroscedastic variances. See, for example, the models considered by Jiang and Zhang (2001). For the case with two heteroscedastic groups of families/clusters, this type of models for the familial count data may be expressed as

$$E[Y_{ij}|\gamma_{i1}^*, \gamma_{i2}^*] = \mu_{ij}^{**} = \exp(\eta_{ij}^*), \tag{4.91}$$

with

$$\eta_{ij}^* = \begin{cases} h(x_{ij}'\beta + \sigma_{\gamma_1}\gamma_{i1}^*), & \text{for } i = 1, \ldots, K_1 \\ h(x_{ij}'\beta + \sigma_{\gamma_2}\gamma_{i2}^*), & \text{for } i = K_1 + 1, \ldots, K, \end{cases}$$

where for both $u=1$ and $u=2$, $\gamma^*_{iu} \overset{iid}{\sim} N(0,1)$ and $\gamma^*_{i1}(i=1,\ldots,K_1)$ and $\gamma^*_{i2}(i=K_1+1,\ldots,K)$ are independent. For the estimation of β, σ_{γ_1}, and σ_{γ_2} parameters, Jiang and Zhang (2001) suggested an improved method of moments (IMM), but as shown by Sutradhar (2004), this IMM approach has several pitfalls. Furthermore, the GQL approach outperforms the so-called IMM approach with regard to the efficiency of the estimators.

Note that the difference between the model (4.89) and (4.91) is that y_{ij}, the count response of the jth member of the ith family, is influenced by the unobservable ith family effect as well as the jth individual effect under the model (4.89), whereas under the model (4.91) y_{ij} is influenced either by the family effects with variance $\sigma^2_{\gamma_1}$ or by the family effects with variance $\sigma^2_{\gamma_2}$.

4.3.3 Multiple Independent Random Effects

In some situations in practice, the count responses under the ith $(i=1,\ldots,K)$ family may be influenced by two or more independent random effects with a distinct component of dispersion. For example, in a clinical study of 'asthma attack' counts for the children of a family, it is reasonable to consider that the frequency of asthma attack on a sibling may be influenced by two random effects components that represent the prevalence of asthma in both the mother's and father's families. Let γ_i and τ_i represent these two unobservable random family effects. Suppose that $\gamma_i \sim \mathrm{N}(0,\sigma^2_\gamma)$ and $\tau_i \sim \mathrm{N}(0,\sigma^2_\tau)$ and γ_i and τ_i are independent for all $i=1,\ldots,K$. Then, unlike (4.89) – (4.90) and (4.91), the conditional mean of the count response y_{ij} now may be written as

$$\mu^*_{ij} = E[Y_{ij}|\gamma_i,\tau_i] = \exp(\eta_{ij}) \tag{4.92}$$

with

$$\eta_{ij} = h(x'_{ij}\beta + z_{i1}\sigma_\gamma\gamma^*_i + z_{i2}\sigma_\tau\tau^*_i), \tag{4.93}$$

where $\gamma^*_i = \gamma_i/\sigma_\gamma$ and $\tau^*_i = \tau^*_i/\sigma_\tau$. In (4.93) β is a $p\times 1$ vector of regression effects and z_{i1} and z_{i2} are known covariates corresponding to γ_i and τ_i. Note that if these covariates are identical, that is, $z_{i1} = z_{i2}$ for all $i=1,\ldots,K$, and $\sigma_\gamma = \sigma_\tau$, then there will be a problem of identification between σ_γ and σ_τ. Thus, in any demonstrations for the effectiveness of any estimation method for the estimation of both variance components, it would be appropriate to consider different values for these variance components in a situation when $z_{i1} = z_{i2}$.

Further note that because the exact likelihood approach, as demonstrated in Section 4.2.1, is complex for a single random effects based Poisson mixed model, this approach will be much more complicated in the present case with two or more variance components. Also, because of the difficulties encountered by the PQL and HL approaches in producing consistent estimates of the variance components, in this section we concentrate on the GQL [Sutradhar (2004)] and MM [Jiang (1998)] approaches only, and for convenience of practitioners, provide the necessary formulas for the construction of the estimating equations under these approaches. These formulas are also available from Jowaheer, Sutradhar, and Sneddon (2009).

4.3.3.1 Method of Moments Estimation for β, σ^2_γ, and σ^2_τ

In this approach, following Jiang (1998) [see also Jiang and Zhang (2001)], one estimates the parameters of the model in (4.92) – (4.93), namely, β, σ^2_γ, and σ^2_τ by using three moment estimating equations that are constructed based on three basic statistics:

$$W_1 = \sum_{i=1}^{K}\sum_{j=1}^{n_i} x_{ij}y_{ij}, \qquad W_2 = \sum_{i=1}^{K} z_{i1}l_i, \qquad W_3 = \sum_{i=1}^{K} z_{i2}l_i \tag{4.94}$$

where $l_i = \sum_{j=1}^{n_i} y^2_{ij} + 2\sum_{j<k} y_{ij}y_{ik}$. Let $\mathbf{w} = (W_1', W_2, W_3)'$ be the $(p+2)$-dimensional vector of these statistics and $\xi = (\xi_1', \xi_2, \xi_3)' = \mathrm{E}(\mathbf{w})$. Similar to (4.57), the moment estimates of the parameters (*i.e.*, of $\theta = (\beta', \sigma^2_\gamma, \sigma^2_\tau)'$), are obtained by solving the estimating equation

$$\mathbf{w} - \xi = 0. \tag{4.95}$$

Let $\hat{\theta}_{MM} = (\hat{\beta}'_{MM}, \hat{\sigma}^2_\gamma, \hat{\sigma}^2_\tau)'$ denote the moment estimator of θ which is the solution of (4.95). This solution may be obtained iteratively by using the customary Newton-Raphson iterative equation

$$\hat{\theta}_{MM}(r+1) = \hat{\theta}_{MM}(r) + (\mathbf{P}')^{-1}_{(r)}(\mathbf{w} - \xi)_{(r)} \tag{4.96}$$

[see also (4.58)], where $(\ )_{(r)}$ denotes the expression within brackets is evaluated at $\hat{\theta}_{MM}(r)$. In (4.96), $\mathbf{P}$ is the $(p+2)\times(p+2)$ derivative matrix of ξ with respect to θ; that is

$$\mathbf{P}' = \begin{pmatrix} \frac{\partial \xi_1}{\partial \beta'} & \frac{\partial \xi_2}{\partial \beta'} & \frac{\partial \xi_3}{\partial \beta'} \\ \\ \frac{\partial \xi_1}{\partial \sigma^2_\gamma} & \frac{\partial \xi_2}{\partial \sigma^2_\gamma} & \frac{\partial \xi_3}{\partial \sigma^2_\gamma} \\ \\ \frac{\partial \xi_1}{\partial \sigma^2_\tau} & \frac{\partial \xi_2}{\partial \sigma^2_\tau} & \frac{\partial \xi_3}{\partial \sigma^2_\tau} \end{pmatrix}. \tag{4.97}$$

Note that the formulas for ξ_1, ξ_2, and ξ_3 may be obtained using the fact that, conditional on γ_i^* and τ_i^*, $y_{ij} \sim \mathrm{Poisson}(\mu^*_{ij})$ with $\mu^*_{ij} = \exp(x'_{ij}\beta + z_{i1}\sigma_\gamma\gamma_i^* + z_{i2}\sigma_\tau\tau_i^*)$ as in (4.92) – (4.93) with identity link function (i.e., $h(g) = g$), and also, y_{ij} and y_{ik} are independent for $j \neq k$, $j,k = 1,\ldots,n_i$. See exercise 4.3 for these formulas. Also, see Exercise 4.4 for the formulas for the derivatives required to construct the derivative matrix $\mathbf{P}$ in (4.97).

4.3.3.2 Joint GQL Estimation for β, σ_γ^2, and σ_τ^2

Recall from Section 4.2.6.3 that for the Poisson mixed model with one variance component σ_γ^2, the components of $\theta = (\beta', \sigma_\gamma^2)'$ were estimated by solving the joint GQL estimating equation (4.85). We still can use this equation given by

$$\sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \theta} \Upsilon_i^{-1}(s_i - \zeta_i) = 0, \tag{4.98}$$

with a difference that θ now has one more component, that is, $\theta = (\beta', \sigma_\gamma^2, \sigma_\tau^2)'$. Also, the formulas for ζ_i and Υ_i in (4.98) will be similar but different from those in (4.85). Once these formulas are known, (4.98) may be solved by using an iterative equation similar to that of (4.86). To compute the elements of ζ_i and Υ_i, we re-express their formulas from (4.80) – (4.82) as

$$\zeta_i = E[S_i] = (\mu_i', \lambda_{i1}', \lambda_{i2}')', \tag{4.99}$$

with

$$\begin{aligned} \mu_i &= [\mu_{i1}, \ldots, \mu_{ij}, \ldots, \mu_{in_i}]' \\ \lambda_{i1} &= [\lambda_{i11}, \ldots, \lambda_{ijj}, \ldots, \lambda_{in_in_i}]' \\ \lambda_{i2} &= [\lambda_{i12}, \ldots, \lambda_{ijk}, \ldots, \lambda_{i(n_i-1)n_i}]', \end{aligned} \tag{4.100}$$

and

$$\begin{aligned} \Upsilon_i = \text{cov}(S_i) &= \begin{bmatrix} \text{cov}(Y_i) & \text{cov}(Y_i, U_{i1}') & \text{cov}(Y_i, U_{i2}') \\ & \text{cov}(U_{i1}) & \text{cov}(U_{i1}, U_{i2}') \\ & & \text{cov}(U_{i2}) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_i & B_i & E_i \\ & \Omega_{i11} & \Omega_{i12} \\ & & \Omega_{i22} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_i & B_i & E_i \\ & F_i & G_i \\ & & H_i \end{bmatrix}, \end{aligned} \tag{4.101}$$

Elements of the ζ_i Vector

The formulas for the general elements in (4.100) are (see Exercise 4.3) given by

$$\mu_{ij} = m_{ij} p_i q_i, \quad \lambda_{ijj} = m_{ij} p_i q_i (1 + m_{ij} p_i^3 q_i^3), \quad \lambda_{ijk} = m_{ij} m_{ik} p_i^4 q_i^4, \tag{4.102}$$

where

$$m_{ij} = \exp(x'_{ij}\beta,\ p_i = \exp(z_{i1}^2\sigma_\gamma^2/2),\ \text{and}\ q_i = \exp(z_{i2}^2\sigma_\tau^2/2).$$

Elements of the Σ_i Matrix

It follows from (4.102) that

$$\begin{aligned}\text{Var}(Y_{ij}) &= \sigma_{ijj} = \mu_{ij}(1-\mu_{ij}+\mu_{ij}p_i^2q_i^2)\\ \text{Cov}(Y_{ij},Y_{ik}) &= \sigma_{ijk} = \lambda_{ijk}-\mu_{ij}\mu_{ik}.\end{aligned}$$

Next, for $\mu_{ij}^* = m_{ij}a_ib_i \equiv m_{ij}\exp(z_{i1}\sigma_\gamma\gamma_i^*)\exp(z_{i2}\sigma_\tau\tau_i^*)$, by using the Poisson based third− and fourth-order conditional moments

$$\begin{aligned}\text{E}(Y_{ij}^3|\gamma_i,\tau_i) &= (\mu_{ij}^*)^3+3(\mu_{ij}^*)^2+\mu_{ij}^*\\ \text{E}(Y_{ij}^4|\gamma_i,\tau_i) &= (\mu_{ij}^*)^4+7(\mu_{ij}^*)^3+6(\mu_{ij}^*)^2+\mu_{ij}^*\end{aligned}$$

and the fact that the Y_{ij} and Y_{ik} are independent for $j \neq k$, conditional on γ_i and τ_i, one may derive the formulas for the elements of all submatrices in (4.101). More specifically:

Elements of the B_i Matrix

$$\text{Cov}(Y_{ij},Y_{ij}^2) = \mu_{ij}[1+3\mu_{ij}p_i^2q_i^2+\mu_{ij}^2p_i^6q_i^6-\mu_{ij}(1+3\mu_{ij}p_i^2q_i^2)] \quad (4.103)$$

$$\text{Cov}(Y_{ij},Y_{ik}^2) = \mu_{ij}\mu_{ik}[(p_i^2q_i^2-1)+\mu_{ik}p_i^2q_i^2(p_i^4q_i^4-1)], \qquad j\neq k \quad (4.104)$$

Elements of the E_i Matrix

$$\begin{aligned}\text{Cov}(Y_{ij},Y_{ij}Y_{il}) &= \mu_{ij}\mu_{il}p_i^2q_i^2(1+\mu_{ij}(p_i^4q_i^4-1)] \qquad j<l \quad (4.105)\\ &= \text{Cov}(Y_{ij},Y_{il}Y_{ij}), \qquad j>l\end{aligned}$$

$$\text{Cov}(Y_{ij},Y_{ik}Y_{il}) = \mu_{ij}\mu_{ik}\mu_{il}p_i^2q_i^2(p_i^4q_i^4-1), \qquad j\neq k\neq l, k<l \quad (4.106)$$

Elements of the F_i Matrix

$$\begin{aligned}\text{Var}(Y_{ij}^2) &= \mu_{ij}[(1+7\mu_{ij}p_i^2q_i^2+6\mu_{ij}^2p_i^6q_i^6+\mu_{ij}^3p_i^{12}q_i^{12})\\ &\quad -\mu_{ij}(1+\mu_{ij}p_i^2q_i^2)^2)] \qquad (4.107)\end{aligned}$$

$$\begin{aligned}\text{Cov}(Y_{ij}^2,Y_{ik}^2) &= \mu_{ij}\mu_{ik}[p_i^2q_i^2(1+(\mu_{ij}+\mu_{ik})p_i^4q_i^4+\mu_{ij}\mu_{ik}p_i^{10}q_i^{10})\\ &\quad -(1+\mu_{ij}p_i^2q_i^2)(1+\mu_{ik}p_i^2q_i^2)], \qquad j\neq k \qquad (4.108)\end{aligned}$$

Elements of the G_i Matrix

$$\begin{aligned}\text{Cov}(Y_{ij}^2, Y_{ij}Y_{ik}) &= \mu_{ij}\mu_{ik}p_i^2q_i^2[(1+3\mu_{ij}p_i^4q_i^4+\mu_{ij}^2p_i^{10}q_i^{10}) \\ &\quad -\mu_{ij}(1+\mu_{ij}p_i^2q_i^2)], \qquad j<k \qquad (4.109)\\ &= \text{Cov}(Y_{ij}^2, Y_{ik}Y_{ij}), \qquad k<j \\ \text{Cov}(Y_{ij}^2, Y_{ik}Y_{il}) &= \mu_{ij}\mu_{ik}\mu_{il}p_i^2q_i^2[p_i^4q_i^4(1+\mu_{ij}p_i^6q_i^6) \\ &\quad -(1+\mu_{ij}p_i^2q_i^2)], \qquad j\neq k\neq l, k<l \qquad (4.110)\end{aligned}$$

Elements of the H_i Matrix

$$\begin{aligned}\text{Var}(Y_{ij}Y_{ik}) &= \mu_{ij}\mu_{ik}p_i^2q_i^2[1+(\mu_{ij}+\mu_{ik})p_i^4q_i^4 \\ &\quad +\mu_{ij}\mu_{ik}p_i^2q_i^2(p_i^8q_i^8-1)], \qquad j<k \qquad (4.111)\\ \text{Cov}(Y_{ij}Y_{ik}, Y_{ij}Y_{im}) &= \mu_{ij}\mu_{ik}\mu_{im}p_i^4q_i^4[p_i^2q_i^2+\mu_{ij}(p_i^8q_i^8-1)], \\ &\quad j<k, j<m, k\neq m \qquad (4.112)\\ &= \text{Cov}(Y_{ij}Y_{ik}, Y_{im}Y_{ij}), \qquad j<k, m<j, k\neq m \\ &= \text{Cov}(Y_{ik}Y_{ij}, Y_{ij}Y_{im}), \qquad k<j, j<m, k\neq m \\ &= \text{Cov}(Y_{ik}Y_{ij}, Y_{im}Y_{ij}), \qquad k<j, m<j, k\neq m \\ \text{Cov}(Y_{ij}Y_{ik}, Y_{il}Y_{im}) &= \mu_{ij}\mu_{ik}\mu_{il}\mu_{im}p_i^4q_i^4 \\ &\quad \times(p_i^8q_i^8-1), \qquad j<k, l<m, j\neq l, k\neq m \quad (4.113)\end{aligned}$$

Note that the construction of the GQL estimating equation (4.98) also requires the formulas for the elements of the $(p+2)\times[n_i(n_i+3)/2]$ derivative matrix $\partial\zeta_i'/\partial\theta$, with $\theta=(\beta,\sigma_\gamma^2,\sigma_\tau^2)'$. Further note that to compute this derivative matrix, it is sufficient to derive the formulas for $\partial\mu_{ij}/\partial\beta$, $\partial\lambda_{ijj}/\partial\beta$, $\partial\lambda_{ijk}/\partial\beta$, $\partial\mu_{ij}/\partial\sigma_\gamma^2$, $\partial\lambda_{ijj}/\partial\sigma_\gamma^2$, $\partial\lambda_{ijk}/\partial\sigma_\gamma^2$, $\partial\mu_{ij}/\partial\sigma_\tau^2$, $\partial\lambda_{ijj}/\partial\sigma_\tau^2$, and $\partial\lambda_{ijk}/\partial\sigma_\tau^2$. These formulas are available from Exercise 4.5.

We are now ready to solve the GQL estimating equation (4.98) for

$$\theta=(\beta',\sigma_\gamma^2,\sigma_\tau^2)'.$$

Let $\hat{\theta}_{GQL}=(\hat{\beta}'_{GQL},\hat{\sigma}^2_{\gamma(GQL)},\hat{\sigma}^2_{\tau(GQL)})'$ be the solution of the GQL estimating equation (4.98). This solution may be achieved by using the iterative equation (4.86), where now ζ_i and Υ_i are given by (4.99) and (4.101), and the formula for the derivative matrix $\partial\zeta_i'/\partial\theta$ is given in Exercise (4.5).

4.3.3.3 Relative Performances of the GQL Versus MM Approaches: An Asymptotic Efficiency Comparison

(a) Asymptotic Variance of the MM Estimator

Note that it follows from (4.96) that $\hat{\theta}_{MM}$ has the asymptotic $(K\to\infty)$ covariance matrix given by

$$\mathrm{Cov}(\hat{\theta}_{MM}) = \mathbf{P}^{-1}\mathbf{V}(\mathbf{P}')^{-1}, \tag{4.114}$$

where $\mathbf{P}$ is the first-order derivative matrix given as in (4.97), and $\mathbf{V}$ is the covariance matrix of $\mathbf{w}$.

For convenience of computation, we express the $\mathbf{V}$ matrix as

$$\mathbf{V} = \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ & V_{22} & V_{23} \\ & & V_{33} \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(W_1) & \mathrm{Cov}(W_1,W_2) & \mathrm{Cov}(W_1,W_3) \\ & \mathrm{Var}(W_2) & \mathrm{Cov}(W_2,W_3) \\ & & \mathrm{Var}(W_3) \end{pmatrix}. \tag{4.115}$$

Note that the elements of the $\mathbf{V}$ matrix in (4.115) can be computed as follows. Recall W_1, W_2, and W_3 as defined in (4.95). It follows that $\mathrm{Var}(W_1)$ is given by

$$\mathrm{Var}(W_1) = \sum_{i=1}^{K} \left[\sigma_{ijj}\mathbf{x}_{ij}\mathbf{x}'_{ij} + 2\sum_{j<k} \sigma_{ijk}\mathbf{x}_{ij}\mathbf{x}'_{ik} \right],$$

where σ_{ijj} and σ_{ijk} are defined following (4.102). Similarly, by exploiting the variances and the covariances from (4.103) to (4.113), the remaining elements of the $\mathbf{V}$ matrix may be computed as follows.

Formula for Cov(W_1,W_2) **:**

$$\begin{aligned}
\mathrm{Cov}(W_1,W_2) &= \sum_{i=1}^{K} \left[\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} z_{i1}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{ik}^2) + 2\sum_{j,k}\sum_{j<l}^{n_i} z_{i1}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{ik}Y_{il}) \right] \\
&= \sum_{i=1}^{K} \left[\sum_{j=1}^{n_i} z_{i1}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{ij}^2) + \sum_{j\neq k} z_{i1}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{ik}^2) \right. \\
&\quad + 2\left(\sum_{j<l} z_{i1}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{ij}Y_{il}) + \sum_{l<j} z_{i1}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{il}Y_{ij}) \right. \\
&\quad \left.\left. + \sum_{j\neq k\neq l,j<l} z_{i1}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{ik}Y_{il}) \right) \right]
\end{aligned} \tag{4.116}$$

Formula for Cov(W_1,W_3)

$$\mathrm{Cov}(W_1,W_3) = \sum_{i=1}^{K} \left[\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} z_{i2}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{ik}^2) + 2\sum_{j,k}\sum_{j<l}^{n_i} z_{i2}\mathbf{x}_{ij}\mathrm{Cov}(Y_{ij},Y_{ik}Y_{il}) \right]$$

Formula for Var(W_2)

$$\mathrm{Var}(W_2) = \mathrm{Var}\left[\sum_{i=1}^{K} z_{i1}^2 \left(\sum_{j=1}^{n_i} Y_{ij}^2 + 2\sum_{j<k}^{n_i} Y_{ij}Y_{ik} \right) \right]$$

$$
\begin{aligned}
&\sum_i z_{i1}^2 \left(\sum_j \text{Var}(Y_{ij}^2) + \sum_{j \neq k} \text{Cov}(Y_{ij}, Y_{ik}) \right) \\
&+ 4 \sum_i z_{i1}^2 \left(\sum_{j<k} \text{Var}(Y_{ij}Y_{ik}) + \sum_{j<k, j<m, k \neq m} \text{Cov}(Y_{ij}Y_{ik}, Y_{ij}Y_{im}) \right. \\
&+ \sum_{j<k, m<j, k \neq m} \text{Cov}(Y_{ij}Y_{ik}, Y_{im}Y_{ij}) + \sum_{k<j, j<m, k \neq m} \text{Cov}(Y_{ik}Y_{ij}, Y_{ij}Y_{im}) \\
&+ \left. \sum_{k<j, m<j, k \neq m} \text{Cov}(Y_{ik}Y_{ij}, Y_{im}Y_{ij}) + \sum_{j<k, l<m, j \neq l, k \neq m} \text{Cov}(Y_{ij}Y_{ik}, Y_{il}Y_{im}) \right) \\
&+ 4 \sum_i z_{i1}^2 \left(\sum_{j<k} \text{Cov}(Y_{ij}^2, Y_{ij}Y_{ik}) + \sum_{k<j} \text{Cov}(Y_{ij}^2, Y_{ik}Y_{ij}) \right. \\
&+ \left. \sum_{j \neq k \neq l, k<l} \text{Cov}(Y_{ij}^2, Y_{ik}Y_{il}) \right). \qquad (4.117)
\end{aligned}
$$

Note that Var(W_3) can be computed from (4.117) by replacing z_{i1}^2 with $z_{i1}z_{i2}$. This completes the computation of the asymptotic covariance in (4.114).

(b) Asymptotic Variance of the GQL Estimator

Note that, unlike the computation of the asymptotic variance of the moment estimator by (4.114), the computation of the asymptotic variance of the GQL estimator is simpler. This is because it follows from (4.98) that, as $K \to \infty$, one obtains the asymptotic covariance of $\hat{\theta}_{GQL}$ as

$$
\text{Cov}(\hat{\theta}_{GQL}) = \lim_{K \to \infty} \left[\sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \theta} \Upsilon_i^{-1} \frac{\partial \zeta_i}{\partial \theta'} \right]^{-1}, \qquad (4.118)
$$

which does not require the computation of any further formulas than those used in (4.98). To be specific, to compute the covariance in (4.118), the formulas for Υ_i and the derivative matrix $\partial \zeta_i'/\partial \theta$ are available from (4.101) and Exercise 4.4, respectively.

(c) Asymptotic Efficiency Comparison: An Illustration

Consider a simple case where all members in a family have the same covariate information; that is, $\mathbf{x}_{ij} = x_i$ with $p = 1$ for all $j = 1, \ldots, n_i$. Let the scalar regression coefficient be denoted by β_1. It then follows that y_{ij}, conditional on γ_i and τ_i, has the Poisson distribution with mean parameter

$$
\exp(x_i\beta_1 + z_{i1}\sigma_\gamma\gamma_i + z_{i2}\sigma_\tau\tau_i).
$$

We consider $K = 100$ and $n_i = 4$ for all $i = 1,\ldots,K$. As far as the fixed effect covariate is concerned, we consider

$$x_i = \begin{cases} -1 & i = 1,\ldots,25 \\ 0 & i = 26,\ldots,75 \\ 1 & i = 76,\ldots,100. \end{cases}$$

For random effects covariates we consider

$$z_{i1} = \begin{cases} 1 & i = 1,\ldots,50 \\ 0 & \text{otherwise} \end{cases} \qquad z_{i2} = \begin{cases} 1 & i = 1,\ldots,25 \\ 0 & \text{otherwise.} \end{cases}$$

Now for selected values of the parameters, namely, $\beta_1 = 1.0$, $\sigma_\gamma = 0.25$, and $\sigma_\tau = 1.0$, we compute the 3×3 covariance matrices $\text{Cov}(\hat{\theta}_{MM})$ by (4.114) and $\text{Cov}(\hat{\theta}_{GQL})$ by (4.118), where $\theta = (\beta_1, \sigma^2_\gamma, \sigma^2_\tau)$. The efficiency of the GQL estimators relative to the moment estimators are found to be

$$\text{eff}(\hat{\beta}_{1,GQL}) = \frac{\text{Var}(\hat{\beta}_{1,MM})}{\text{Var}(\hat{\beta}_{1,GQL})} = 2.06, \qquad \text{eff}(\hat{\sigma}^2_{\gamma,GQL}) = \frac{\text{Var}(\hat{\sigma}^2_{\gamma,MM})}{\text{Var}(\hat{\sigma}^2_{\gamma,GQL})} = 2.40,$$

$$\text{eff}(\hat{\sigma}^2_{\tau,GQL}) = \frac{\text{Var}(\hat{\sigma}^2_{\tau,MM})}{\text{Var}(\hat{\sigma}^2_{\tau,GQL})} = 2.13,$$

showing that the GQL estimator is 2.06 times more efficient in estimating β_1, 2.40 times more efficient in estimating σ^2_γ, and 2.13 times more efficient in estimating σ^2_τ than the MM approach.

4.3.3.4 GQL Versus MM Estimation: A Simulation Study Based on an Asthma Count Data Model with Two Components of Dispersion

4.3.3.5 An Asthma Count Data Model with Four Fixed Covariates and Two Components of Dispersion

The purpose of this section is to examine the relative performances of the MM and GQL estimation approaches for the estimation of the parameters of a Poisson mixed model with random effects from two independent sources. We consider hypothetical asthma data where any asthma attack on a sibling may be influenced by two random effects components that represent the prevalence of asthma in both the mother's and father's families. Jowaheer, Sutradhar and Sneddon (2009) have conducted a simulation study to examine the performances of the MM and GQL approaches for this type of asthma data. For convenience, we explain their simulation results here.

Along with two sources of random effects, it is assumed that the mean of the Poisson data are affected by four fixed covariates, namely, child's gender (x_{ij1}), child's age (x_{ij2}), mother's smoking habit (x_{ij3}), and father's smoking habit (x_{ij4}). Following (4.92), we now write the model with Poisson mean

$$\mu_{ij}^* = \exp(x_{ij1}\beta_1 + x_{ij2}\beta_2 + x_{ij3}\beta_3 + x_{ij4}\beta_4 + z_{i1}\sigma_\gamma\gamma_i^* + z_{i2}\sigma_\tau\tau_i^*). \quad (4.119)$$

Assume that there is a 20% chance of asthma being present in the mother's side. Therefore, we use $z_{i1} \sim \text{Bin}(0.2)$. Similarly, we assume that there is a 10% chance of asthma in the father's side. Therefore, we use $z_{i2} \sim \text{Bin}(0.1)$. Next suppose that there is a 50% probability for a child being male. Therefore $x_{ij1} \sim \text{Bin}(0.5)$. We further assume that the age is uniformly distributed between 1 and 20, denoted as Unif(1, 20), then create $x_{ij2} = \text{Unif}(1,20)/100$. Next we consider that there is a 25% chance that the mother smokes, and 35% chance that the father smokes. Therefore

$$x_{ij3} \sim \text{Bin}(0.25) \text{ and } x_{ij4} \sim \text{Bin}(0.35)$$

.

With regard to the model parameters, we consider 100 families (*i.e.*, $K = 100$), each with size 4 (*i.e.*, $n_i = n = 4$). As far as the effects of the four fixed covariates are concerned, we use $\beta_1 = 0.5$ (gender effect), $\beta_2 = 0.05$ (age effect of the child), $\beta_3 = 0.3$ (effect of mother's smoking status) and $\beta_4 = 0.1$ (effect of father's smoking status). For the random hereditary effects, we use $\sigma_\gamma^2 = 1$ for the mother's side and $\sigma_\tau^2 = 0.25$ and 0.5625 for the father's side. These parameters were estimated based on 500 simulations by solving the moment equations (4.95) under the MM approach and by solving the GQL estimating equation (4.98) under the GQL approach. The simulation results are shown in Table 4.11. For similar simulated results with more model parameter we refer to Jowaheer et al (2009).

Table 4.11 The joint MM and joint GQL Estimates along with the corresponding estimated standard errors and mean squared errors for the hypothetical asthma data generated following the Poisson mixed model (4.119) with random effects from two independent sources.

σ_γ^2	σ_τ^2	Method	Quantity	$\hat\beta_1$	$\hat\beta_2$	$\hat\beta_3$	$\hat\beta_4$	$\hat\sigma_\gamma^2$	$\hat\sigma_\tau^2$
1.0	0.25	GQL	Mean	0.496	0.074	0.296	0.094	0.905	0.186
			SE	0.080	0.563	0.087	0.077	0.431	0.212
			MSE	0.006	0.316	0.007	0.006	0.194	0.049
		MM	Mean	0.506	0.204	0.299	0.098	0.729	0.055
			SE	0.092	0.686	0.124	0.097	0.325	0.419
			MSE	0.009	0.493	0.015	0.009	0.179	0.213
	0.5625	GQL	Mean	0.501	0.048	0.295	0.096	0.853	0.455
			SE	0.075	0.546	0.086	0.082	0.423	0.408
			MSE	0.006	0.297	0.007	0.007	0.200	0.178
		MM	Mean	0.513	0.129	0.306	0.100	0.689	0.239
			SE	0.098	0.693	0.116	0.119	0.325	0.459
			MSE	0.010	0.485	0.013	0.014	0.202	0.315

It is interesting to note that both the GQL and MM approaches estimate the regression effect well, with GQL being more efficient. This is because the MSE of the GQL regression estimates appears to be the same or always smaller than the moment estimates. The performance of the GQL estimates is much better than the moment estimates when σ_τ^2 is large. As far as the estimates of the variance components (reflecting hereditary differences) are concerned, GQL provides better estimates than the moment approach in estimating σ_τ^2, whereas the MM approach appears to perform slightly better than the GQL approach in estimating the large value of $\sigma_\gamma^2 = 1.0$. However, the results in Jowaheer, Sutradhar, and Sneddon (2009, Table 6) show that for almost all values of σ_γ^2 and σ_τ^2, the GQL approach performs better than the MM approach in estimating these variance parameters. In summary, the GQL approach, as compared to the MM approach, yields better (consistent and efficient) estimates for all parameters including the regression effects involved in the models with two variance components.

4.4 Semiparametric Approach

For simplicity, consider the single random effects case, and similar to (4.93), write

$$\eta_{ij} = \theta_{ij} + z_{i1}\gamma_i, \text{ with } \theta_{ij} = x'_{ij}\beta.$$

Also, for convenience, write the Poisson density from (4.1) in the exponential form as

$$f(y_{ij}|\eta_{ij}) = \exp[\{y_{ij}\eta_{ij} - a(\eta_{ij})\} + b(y_{ij})], \tag{4.120}$$

where $a(\eta_{ij}) = \exp(\eta_{ij})$ and $b(\cdot)$ is a known function free from parameters. Note that even though it is practically an appealing assumption that $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma_\gamma^2)$, some authors have studied the inferences in the generalized linear mixed model setup, by relaxing the distributional assumptions for the random effects γ_i. For example, instead of the normal distribution assumption, Sutradhar and Rao (2001) have assumed that the moments of γ_i up to order four are known and the higher moments of order more than four are negligible. Thus, they have considered a semiparametric model, where it is known that conditional on γ_i, the responses follow the Poisson distribution, but, the distribution of γ_i is not known. To be specific, for moments of the random effects γ_i, they assumed that

$$E\gamma_i^r = \delta_r(\sigma_\gamma^2) = \sum_{s=1}^{r} c_{r,s}\sigma_\gamma^{r+1-s}, \quad \text{for } r = 1, \ldots, 4, \tag{4.121}$$

and

$$E\gamma_i^r = o(\sigma_\gamma^r), \quad \text{for } r \geq 5,$$

where $c_{r,s}$ are suitable known constants for $r = 1, \ldots, 4$. For example, if $\gamma_i \sim N(0, \sigma_\gamma^2)$, then $c_{1,1} = 0$, $c_{2,1} = 1$, $c_{2,2} = 0$, $c_{3,1} = c_{3,2} = c_{3,3} = 0$, and $c_{4,1} = 3$, $c_{4,2} = c_{4,3} = c_{4,4} = 0$.

Note that for the estimation of β and σ_γ^2, Sutradhar and Rao (2001) have used the marginal QL (MQL) approach, which we have referred to as the marginal GQL approach in Section 4.2.6. More specifically, the regression effects β and the variance component σ^2_γ, may be estimated by solving the marginal GQL estimating equations (4.62) and (4.69), respectively. However, these estimating equations are now constructed by using the model (4.120) so that the moment assumptions in (4.121) are satisfied. For convenience, we re-express the estimating equations (4.62) for β, and (4.69) for σ^2_γ, as follows.

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1}(y_i - \mu_i) = 0, \quad \sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma^2_\gamma} \Omega_i^{-1}(u_i - \lambda_i) = 0,$$

where for $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$,

$$\mu_i = E[Y_i], \text{ and } \Sigma_i = \text{var}[Y_i],$$

and for $u_i = (u_{i1}', u_{i2}')'$, with

$$\begin{aligned} u_{i1} &= (y^2_{i1}, \ldots, y^2_{ij}, \ldots, y^2_{in_i})' : n_i \times 1, \\ u_{i2} &= (y_{i1}y_{i2}, \ldots, y_{ij}y_{ik}, \ldots, y_{i(n_i-1)}y_{in_i})', j < k : \frac{n_i(n_i-1)}{2} \times 1, \end{aligned}$$

$$\lambda_i = E[U_i] \text{ and } \Omega_i = \text{var}[U_i].$$

Now to compute

$$\mu_i, \ \lambda_i, \ \Sigma_i, \text{ and } \Omega_i,$$

such that the moment conditions in (4.121) are satisfied, we first develop an approximation to the unconditional joint distribution of the responses $y_{i1}, \ldots, y_{in_i}$, satisfying (4.121). For convenience, we provide this approximation in the likelihood form as

$$\begin{aligned} L_i(\beta, \sigma^2_\gamma) = & \left[\Pi_{j=1}^{n_i} f(y_{ij}|\theta_{ij})\right] \left[1 + \left[\frac{\sigma^2_\gamma}{2}\{A_i^2 - B_i\} + \frac{\delta_3(\sigma^2_\gamma)}{6}\{A_i^3 - 3A_iB_i - C_i\}\right.\right. \\ & \left.\left. + \frac{\delta_4(\sigma^2_\gamma)}{24}\{A_i^4 - 6A_i^2B_i - 4A_iC_i + 3B_i^2 - D_i\}\right]\right], \end{aligned} \tag{4.122}$$

where $f(y_{ij}|\theta_{ij}) = f(y_{ij}|\eta_{ij})|_{\eta_{ij}=\theta_{ij}}$, $\delta_r(\sigma^2_\gamma)$ for $r = 1, \ldots, 4$, are as in (4.121), and

$$A_i = \sum_{j=1}^{n_i} z_{i1}(y_{ij} - a'_{ij}), \qquad B_i = \sum_{j=1}^{n_i} z^2_{i1} a''_{ij},$$

$$C_i = \sum_{j=1}^{n_i} z^{\frac{3}{2}}_{i1} a'''_{ij}, \quad \text{and} \quad D_i = \sum_{j=1}^{n_i} z^2_{i1} a^{IV}_{ij},$$

with a'_{ij}, a''_{ij}, a'''_{ij}, and a^{IV}_{ij} as the first−, second−, third−, and the fourth-order derivatives of $a(\eta_{ij}) = \exp(\eta_{ij})$ in (4.120) with respect to η_{ij} and then evaluated at $\eta_{ij} = \theta_{ij}$.

To derive the likelihood function in (4.122), we first write

$$L_i(\beta, \sigma^2_\gamma) = \int L^*_i(\beta, \gamma_i) g_N(\gamma_i|\sigma^2_\gamma) d\gamma_i$$

$$= \int \Pi^{n_1}_{j=1} f(y_{ij}|\eta_{ij}) g_N(\gamma_i|\sigma^2_\gamma) d\gamma_i, \tag{4.123}$$

where $f(y_{ij}|\eta_{ij})$ is the Poisson density (4.120) of y_{ij} conditional on γ_i. Next, one may expand the conditional density $f(y_{ij}|\eta_{ij})$ in (4.123) about θ_{ij} and take the expectation over η_{ij} under the assumption that $E(\gamma^r_i) = o(\sigma^r_\gamma)$, for $r \geq 5$. This operation along with $c_{1,1} = 0$, $c_{2,1} = 1$, and $c_{2,2} = 0$, yields the approximate likelihood for the data as in (4.122).

Computation of Univariate and Joint Probability Density Function

One utilizes the n_i-dimensional joint density (4.122) to obtain a proper unconditional density of y_{ij} and the joint density, for example, for y_{ij} and y_{ik}. For convenience, following the notation for the joint density in (4.122), let

$$L_{i,j},\ L_{i,jk},\ L_{i,jk\ell},\ \text{and}\ L_{i,jk\ell m}$$

denote the univariate, bivariate, trivariate, and four-dimensional pdf of

$$y_{ij};\ y_{ij}, y_{ik};\ y_{ij}, y_{ik}, y_{i\ell};\ \text{and}\ y_{ij}, y_{ik}, y_{i\ell}, y_{im},$$

respectively. For example, by integrating over $y_{ij'}$ for all $j' = 1, \ldots, j-1, j+1, \ldots, n_i$, the marginal density of y_{ij} follows from (4.122), and is given by

$$L_{i,j}(y_{ij}) = f(y_{ij}|\theta_{ij}) \left[1 + \left[\frac{\sigma^2_\gamma}{2} \{A^2_{i,j} - B_{i,j}\} + \frac{\delta_3(\sigma^2_\gamma)}{6} \{A^3_{i,j} - 3A_{i,j}B_{i,j} - C_{i,j}\} \right.\right.$$
$$\left.\left. + \frac{\delta^4(\sigma^2_\gamma)}{24} \{A^4_{i,j} - 6A^2_{i,j}B_{i,j} - 4A_{i,j}C_{i,j} \quad + 3B^2_{i,j} - D_{i,j}\} \right] \right], \tag{4.124}$$

where $A_{i,j} = z_{i1}(y_{ij} - a'_{ij})$, $B_{i,j} = z^2_{i1} a''_{ij}$, $C_{i,j} = z^{3/2}_{i1} a'''_{ij}$, and $D_{i,j} = z^2_{i1} a^{IV}_{ij}$. In a manner similar to that of the derivation of the univariate density $L_{i,j}(\cdot)$, we can derive the desired joint densities, namely, $L_{i,jk}$, $L_{i,jk\ell}$, and $L_{i,jk\ell m}$.

4.4.1 Computations for μ_i, λ_i, Σ_i, and Ω_i

Note that the computations for

$$\mu_i,\ \lambda_i,\ \Sigma_i, \text{ and } \Omega_i,$$

will require moments up to order eight. For this purpose, we first provide the first eight moments of y_{ij} when its probability function is given by (4.20) with $\eta_{ij} = \theta_{ij}$, and $a(\theta_{ij}) = \exp(\theta_{ij})$. Let $m_{ij,1} = E_{\exp}(y_{ij})$ and $m_{ij,s} = E_{\exp}(y_{ij} - m_{ij,1})^s$ for $s = 2,\ldots,8$, with pdf of y_{ij} as in (4.120). The formulas for these moments are available from Exercise 4.6.

Furthermore, to derive the moments of the distribution (4.124) of y_{ij} of a finite-order r, say, it is convenient to compute an integral as in the following lemma.

Lemma 4.4. Let $h^{(r)}_{ij,(r+s)}$ denote the integral

$$h^{(r)}_{ij,(r+s)} = \int y^r_{ij} A^s_{i,j} f_{ij}(y_{ij}|\theta_{ij}) dy_{ij}, \tag{4.125}$$

where $A_{i,j} = z_{i1}(y_{ij} - m_{ij,1})$, $f(\cdot)$ is the exponential density as in (4.120), and r and s are nonnegative integers. Then for $r = 1,2$, and $s = 0,1,2,3,4$, the hs are given by

$$h^{(1)}_{ij,1} = m_{ij,1}, \quad h^{(1)}_{ij,2} = z_{i1} m_{ij,2},$$

$$h^{(1)}_{ij,(1+s)} = z^s_{i1}\{m_{ij,(1+s)} + m_{ij,s} m_{ij,1}\}, \text{ for } s = 2,3,4,$$

and

$$h^{(2)}_{ij,2} = m_{ij,2} + (m_{ij,1})^2, \quad h^{(2)}_{ij,3} = z_{i1}\{m_{ij,3} + 2m_{ij,2} m_{ij,1}\},$$

$$h^{(2)}_{ij,(2+s)} = z^s_{i1}\{m_{ij,(2+s)} + 2m_{ij,(1+s)} m_{ij,1} + m_{ij,s}(m_{ij,1})^2\},$$

for $s = 2,3,4$, where $m_{ij,1}, m_{ij,2}, \ldots, m_{ij,6}$ are as in Exercise 4.6.

Computation of $E[Y^r_{ij}]$ from Marginal pdf $L_{i,j}$ (4.124)

Because $E[Y^r_{ij}] = \int Y^r_{ij} L_{i,j}(y_{ij}) dy_{ij}$, where $L_{i,j}(y_{ij})$ (4.124) is the marginal density of y_{ij}, by using the results from Lemma 4.4, one obtains

$$E(Y^r_{ij}) = h^{(r)}_{ij,(r)} + \frac{\sigma^2_\gamma}{2} d^{(r)}_{ij,(r+2)} + \frac{\delta_3(\sigma^2_\gamma)}{2} d^{(r)}_{ij,(r+3)} + \frac{\delta_4(\sigma^2_\gamma)}{24} d^{(r)}_{ij,(r+4)}, \tag{4.126}$$

where

$$d^{(r)}_{ij,(r+2)} = h^{(r)}_{ij,(r+2)} - h^{(r)}_{ij,(r)}, \quad d^{(r)}_{ij,(r+3)} = h^{(r)}_{ij,(r+3)} - 3h^{(r)}_{ij,(r+1)} B_{i\cdot j} - \phi h^{(r)}_{ij,(r)} C_{i\cdot j},$$

and

$$d^{(r)}_{ij,(r+4)} = h^{(r)}_{ij,(r+4)} - 6h^{(r)}_{ij,(r+2)}B_{i,j} - 4h^{(r)}_{ij,(r+1)}C_{i,j} + 3h^{(r)}_{ij,(r)}B^2_{i,j} - h^{(r)}_{ij,(r)}D_{i,j},$$

with $B_{i,j}$, $C_{i,j}$, and $D_{i,j}$ given as in (4.125).

Computation of Product Moments $E[Y^r_{ij}Y^s_{ik}]$ from Bivariate pdf $L_{i,jk}$

Note that by integrating out $n_i - 2$ variables (all variables except y_{ij} and y_{ik}) over the joint density in (4.123), similar to (4.124), one now obtains the bivariate p.d.f. given by

$$\begin{aligned} L_{i,jk}(y_{ij}, y_{ik}) &= f(y_{ij}|\theta_{ij}) f(y_{ik}|\theta_{ik}) \\ &\times \left[1 + \left[\frac{\sigma^2_\gamma}{2}\{A^2_{i,jk} - B_{i,jk}\} + \frac{\delta_3(\sigma^2_\gamma)}{6}\{A^3_{i,jk} - 3A_{i,jk}B_{i,jk} - C_{i,jk}\}\right.\right. \\ &+ \frac{\delta^4(\sigma^2_\gamma)}{24}\{A^4_{i,jk} - 6A^2_{i,jk}B_{i,jk} - 4A_{i,jk}C_{i,jk} \\ &\left.\left. +3B^2_{i,jk} - D_{i,jk}\}\right]\right], \end{aligned} \tag{4.127}$$

where $A_{i,jk} = z_{i1}[(y_{ij} - a'_{ij}) + (y_{ik} - a'_{ik})]$, $B_{i,jk} = z^2_{i1}[a''_{ij} + a'_{ik}]$, $C_{i,jk} = z^{3/2}_{i1}[a'''_{ij} + a''_{ik}]$, and $D_{i,jk} = z^2_{i1}[a^{IV}_{ij} + a^{IV}_{ik}]$.

For the purpose of computing

$$E[Y^r_{ij}Y^s_{ik}] = \int Y^r_{ij}Y^s_{ik}L_{i,jk}(\cdot)dy_{ij}dy_{ik},$$

similar to Lemma 4.4, it is convenient to perform some more basic integrations with respect to the exponential pdf (4.120) as in the following lemma.

Lemma 4.5. For $j \neq k$, let $H^{(r,s)}_{ijk,(r+s+q)}$ denote the integral

$$H^{(r,s)}_{ijk,(r+s+q)} = \int y^r_{ij}y^s_{ik}A^q_{i,jk} f(y_{ij}|\theta_{ij}) f(y_{ik}|\theta_{ik})dy_{ij}dy_{ik}, \tag{4.128}$$

where, $f(y_{ij}|\theta_{ij})$, for example, is the exponential pdf as in (4.120), $A_{i,jk} = z_{i1}[(y_{ij} - m_{ij,1}) + (y_{ik} - m_{ik,1})]$ as in (4.127), and r, s, and q are nonnegative integers. Then, for $r = 1$, $s = 1$, and $q = 0, 1, \ldots, 4$, the H functions are given by

$$H^{(r,s)}_{ijk,(r+s+q)} = \sum_{u=1}^{q+1} {}^tC_{u-1} h^{(r)}_{ij,(t+2-u)} h^{(s)}_{ik,u}, \tag{4.129}$$

where h functions are as in Lemma 4.4, and ${}^tC_{u-1}$ denotes the number of ways that $u - 1$ functions can be chosen from q functions.

The results of Lemma 4.5 are next exploited to derive the desired product moments as

$$E(Y_{ij}^r Y_{ik}^s) = H_{ijk,(r+s)}^{(r,s)} + \frac{\sigma_\gamma^2}{2} d_{ijk,(r+s+2)} + \frac{\delta_3(\sigma_\gamma^2)}{6} d_{ijk,(r+s+3)}^{(r,s)} + \frac{\delta_4(\sigma_\gamma^2)}{24} d_{ijk,(r+s+4)}^{(r,s)}, \tag{4.130}$$

where for $u = 2,3,4$, $d_{ijk,(r+s+u)}^{(r,s)}$ are obtained from $d_{ij,(r+s+u)}^{(r)}$ in (4.126), by replacing $h_{ij,(r+s+u)}^{(r)}$, $B_{i,j}$, and $C_{i,j}$ with $H_{ijk,(r+s+u)}^{(r,s)}$, $B_{i,jk}$ and $C_{i,jk}$ respectively.

Formulas for μ_i, λ_i, and Σ_i

Let $M_{ij,1}$ denote the jth component of the $\mu_i = E[Y_i]$ vector. Similarly, let $M_{ijj,2} = E[Y_{ij}^2]$ and $M_{ijk,2} = E[Y_{ij}Y_{ik}]$ denote the two general elements of the λ_i vector. the formulas for $M_{ij,1}$ and $M_{ijj,2}$ follow from (4.126), and the formula for $M_{ijk,2}$ follows from (4.130), and these formulas are given by

$$M_{ij,1} = a'_{ij} + \frac{\sigma_\gamma^2}{2} z_{i1}^2 a'''_{ij} + \frac{\delta_3(\sigma_\gamma^2)}{6} z_{i1}^3 a_{ij}^{IV} + \frac{\delta_4(\sigma_\gamma^2)}{24} z_{i1}^4 a_{ij}^{V}, \tag{4.131}$$

$$\begin{aligned} M_{ijj,2} = & \Big[(a'_{ij})^2 + \sigma_\gamma^2 z_{i1}^2 \{a'_{ij}a'''_{ij} + (a''_{ij})^2\} + \frac{\delta_3(\sigma_\gamma^2)}{3} z_{i1}^3 \{a'_{ij}a_{ij}^{IV} \\ & + 3a''_{ij}a'''_{ij}\} + \frac{\delta_4(\sigma_\gamma^2)}{12} z_{i1}^4 \{a'_{ij}a_{ij}^{V} + 4a''_{ij}a_{ij}^{IV} + 3(a'''_{ij})^2\} \Big] \\ & + \left[a''_{ij} + \frac{\sigma_\gamma^2}{2} z_{i1}^2 a_{ij}^{IV} + \frac{\delta_3(\sigma_\gamma^2)}{6} z_{i1}^3 a_{ij}^{V} + \frac{\delta_4(\sigma_\gamma^2)}{24} z_{i1}^4 a_{ij}^{VI} \right] \end{aligned} \tag{4.132}$$

$$\begin{aligned} M_{ijk,2} = & a'_{ij}a'_{ik} + \frac{\sigma_\gamma^2}{2} z_{i1}^2 \left[a'''_{ij}a'_{ik} + 2a''_{ij}a''_{ik} + a'_{ij}a'''_{ik}\right] \\ & + \frac{\delta_3(\sigma_\gamma^2)}{6} z_{i1}^3 \left[a_{ij}^{IV}a'_{ik} + 3a'''_{ij}a''_{ik} + 3a''_{ij}a'''_{ik} + a'_{ij}a_{ik}^{IV}\right] \\ & + \frac{\delta_4(\sigma_\gamma^2)}{24} z_{i1}^4 \left[a_{ij}^{V}a'_{ik} + 4a_{ij}^{IV}a''_{ik} + 6a'''_{ij}a'''_{ik} + 4a''_{ij}a_{ik}^{IV} + a'_{ij}a_{ik}^{V}\right]. \end{aligned} \tag{4.133}$$

The diagonal elements of the Σ_i are then computed by

$$\sigma_{i,jj} = M_{ijj,2} - [M_{ij,1}]^2, \tag{4.134}$$

and similarly, the off-diagonal elements are computed by

$$\sigma_{i,jk} = M_{ijk,2} - [M_{ij,1}M_{ik,1}]. \tag{4.135}$$

Formulas for the Elements of Ω_i

For the construction of this matrix, one needs to compute the third− and fourth-order moments, whereas the elements of the Σ_i matrix were computed by using the formulas for the second-order moments from (4.126) and (4.130). Similar to these second order moments, one may compute the third− and fourth-order moments, namely $E[Y_{ij}^r Y_{ik}^s Y_{i\ell}^t]$ and $E[Y_{ij}^r Y_{ik}^s Y_{i\ell}^t Y_{im}^u]$. The derivation of these moments is, however, lengthy and not given here. For the readers interested in these formulas, we refer to Theorems 4 and 5 in Sutradhar and Rao (2001).

4.4.2 Construction of the Estimating Equation for β When σ_γ^2 Is Known

This estimating equation is given by $\sum_{i=1}^K [\partial\mu_i'/\partial\beta]\Sigma_i^{-1}(y_i - \mu_i) = 0$. Note that the formulas for the elements of μ_i are developed as in (4.131), and Σ_i may be computed by (4.134) and (4.135). We are now left with the computation of $\partial\mu_i'/\partial\beta$, where

$$\mu_i = M_{i,1} = [M_{i1,1}, \ldots, M_{ij,1}, \ldots, M_{in_i,1}]',$$

with $M_{ij,1} = E[Y_{ij}]$ as in (4.131). This first-order derivative matrix is given as follows.

Computation of $\frac{\partial M_{i,1}'}{\partial\beta}$

Note that $\partial\{M_{i,1}'\}/\partial\beta$ is the $n_i \times p$ first derivative matrix of $M_{i,1}'$ with respect to β. The formula for this matrix can be derived by computing the derivative of $M_{ij,1}$ in (4.131) with respect to β. Because $\theta_{ij} = x_{ij}'\beta$, and $a_{ij}', \ldots, a_{ij}^V$ are, respectively, the first five order derivatives of $a_{ij} = \exp(\theta_{ij})$ with respect to θ_{ij}, it then follows from (4.131) that

$$\begin{aligned}\partial M_{ij,1}/\partial\beta &= \left[a_{ij}'' + \frac{\sigma_\gamma^2}{2} z_{i1}^2 a_{ij}^{IV} + \frac{\delta^3(\sigma_\gamma^2)}{6} z_{i1}^3 a_{ij}^V + \frac{\delta_4(\sigma_\gamma^2)}{24} z_{i1}^4 a_{ij}^{VI}\right] x_{ij}', \\ &= w_{ij}x_{ij}', \text{ (say)},\end{aligned}$$

so that

$$\partial M'_{i,1}/\partial\beta = \begin{bmatrix} w_{i1}x'_{i1} \\ \vdots \\ w_{in_i}x'_{in_i} \end{bmatrix} = \begin{pmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & w_{in_i} \end{pmatrix} X_i$$

$$= W_i X_i, \text{ (say)},$$

where $W_i = \text{diag}[w_{i1},\ldots,w_{in_i}]$ and $X_i = [x_{i1},\ldots,x_{ij},\ldots,x_{in_i}]'$. This completes the construction of the estimating function $\sum_{i=1}^{K}[\partial\mu'_i/\partial\beta]\Sigma_i^{-1}(y_i-\mu_i)$, for β, under the present semiparametric (SP) model, when σ_γ^2 is known.

Let $\hat{\beta}_{GQL,SP}$ denote the GQL estimate of β obtained by solving the SP model based estimating equation

$$\sum_{i=1}^{K} \frac{\partial M'_{i,1}}{\partial\beta} M_{i,2}^{-1}(y_i - M_{i,1}) = 0,$$

where, for convenience of notation, $M_{i,1}$ is used for μ_i, and $M_{i,2}$ is used for $\Sigma_i = (\sigma_{i,jk})$, with $\sigma_{i,jj}$ and $\sigma_{i,jk}$ as given by (4.134) and (4.135), respectively. Furthermore, similar to (4.64), it can be shown that asymptotically (as $K \to \infty$), for known σ_γ^2, the GQL estimator $\hat{\beta}_{GQL,SP}$ follows the multivariate Gaussian distribution with mean β and the covariance matrix given by

$$\text{cov}(\hat{\beta}_{GQL_SP}) = \text{limit}_{K\to\infty}\left[\sum_{i=1}^{K} \frac{\partial M'_{i,1}}{\partial\beta} M_{i,2}^{-1} \frac{\partial M_{i,1}}{\partial\beta'}\right]^{-1}. \tag{4.136}$$

4.5 Monte Carlo Based Likelihood Estimation

For simplicity in explaining this approach, we consider the Poisson mixed model with a single random effect as in Sections 4.2 and 4.4. Thus, we use the model (4.120) and using slightly different notations, rewrite

$$f(y_{ij}|\gamma_i,\beta) = \exp[\{y_{ij}\eta_{ij} - a(\eta_{ij})\} + b(y_{ij})], \tag{4.137}$$

with $a(\eta_{ij}) = \exp(\eta_{ij})$, where $\eta_{ij} = \theta_{ij} + z_{i1}\gamma_i$, with $\theta_{ij} = x'_{ij}\beta$. As far as the distributional assumption for γ_i is concerned, we use

$$\gamma_i \overset{\text{iid}}{\sim} N(0,\sigma_\gamma^2);\ that is, g_N(\gamma_i|\sigma_\gamma^2) = (2\pi\sigma_\gamma^2)^{-1/2}\ \exp\{-\gamma_i{}^2/2\sigma_\gamma^2\} \tag{4.138}$$

as in Section 4.2. It follows from (4.137) and (4.138) that the likelihood for the data is given by

$$L(\beta,\sigma_\gamma^2) = \int \Pi_{i=1}^{K} f(y_i|\gamma_i) g_N(\gamma_i|\sigma_\gamma^2) d\gamma_i, \tag{4.139}$$

where $f(y_i|\gamma_i) = \Pi_{j=1}^{n_i} f(y_{ij}|\gamma_i)$.

Recall from Section 2.1 where β and σ_γ^2 were estimated by maximizing the exact likelihood constructed by simulating [Jiang (1998)] the random effects. For a similar but different direct simulated likelihood estimation, one may also refer to Geyer and Thompson (1992) and Gelfand and Carlin (1993), for example. In this section, we highlight a different Monte Carlo approach that has been developed based on the so-called Metropolis algorithm. To be specific, the Metropolis algorithm is used to simulate the random effects and the so-called expectation-maximization (EM) or Newton−Ralphson (NR) technique is used to maximize the Monte Carlo (simulated) based approximate likelihood function. One may be referred to McCulloch (1997) for these MCEM and MCNR approaches. For convenience, in explaining these two approaches, we first outline the common Metropolis algorithm to generate random effects, as follows.

Metropolis Algorithm

Recall from (4.18) that a closed-form likelihood function cannot be obtained due to the problem of integration over the distribution of the random effect γ_i. To handle such an integration problem, some numerical algorithms are developed where γ_i is considered to be a missing dataum, and it is drawn from a conditional distribution of $\gamma_i|y$ by using the Metropolis algorithm [Tanner (1993)], which does not require specification of the unconditional density of the data y. To be specific, a candidate distribution $h(\gamma_i)$ is considered, from which potential new values are drawn, and also an acceptance function is considered that gives the probability of accepting the new value. Suppose that γ_i^+ denotes a new value generated from $h(\gamma_i)$, whereas γ_i^- is the previous value drawn from the conditional distribution of $\gamma_i|y$. The new value γ_i^+ is accepted with a probability $A(\gamma_i^-, \gamma_i^+)$, (say); otherwise the previous value γ_i^- is retained. Denote this first-time decided value, whether γ_i^- or γ_i^+, as $\gamma_i^{(1)}$. Continue this operation for a large number of times, say N, and denote these values of random effect as

$$\gamma_i^{(1)}, \ldots, \gamma_i^{(w)}, \ldots, \gamma_i^{(N)}. \tag{4.140}$$

Note that these values are chosen in a different way from the simulated random effect values $\gamma_{i1}^*, \ldots, \gamma_{iw}^*, \ldots, \gamma_{iN}^*$, used in (4.21). The formula for the acceptance probability $A(\gamma_i^-, \gamma_i^+)$ is given by

$$A(\gamma_i^-, \gamma_i^+) = \min\left[1, \frac{f(\gamma_i^+|y, \beta, \sigma_\gamma^2)h(\gamma_i^-)}{f(\gamma_i^-|y, \beta, \sigma_\gamma^2)h(\gamma_i^+)}\right]. \tag{4.141}$$

The random effect values from (4.140) are then exploited to develop MCEM and MCNR [McCulloch (1997, Sections 3.1, 3.2)] as follows:

4.5.1 MCEM Approach

1. Choose starting values $\beta^{(0)}$ and $\sigma_\gamma^{2(0)}$. Set $r = 0$.

2. Generate N values (4.140) of the random effect by using $f(\gamma_i^-|y, \beta^{(r)}, \sigma_\gamma^{2(r)})$, and
(a) Choose $\beta^{(r+1)}$ to maximize a Monte Carlo estimate of $E[\log f(y|\gamma_i, \beta)]$; that is, maximize

$$\frac{1}{/}N \sum_{w=1}^{N} \log f(y|\gamma_i^{(w)}, \beta). \tag{4.142}$$

(b) Choose $\sigma_\gamma^{2(r+1)})$ to maximize $[1/N] \sum_{w=1}^{N} \log g(\gamma_i^{(w)}|\sigma_\gamma^2)$.
(c) Set $r = r + 1$.

3. If convergence is achieved, then declare $\beta^{(r+1)}$ and $\sigma_\gamma^{2(r+1)}$ to be the maximum likelihood estimates; otherwise, go back to Step 2.

4.5.2 MCNR Approach

1. Choose starting values $\beta^{(0)}$ and $\sigma_\gamma^{2(0)}$. Set $r = 0$.

2. Generate N values (4.140) of the random effect by using $f(\gamma_i^-|y, \beta^{(r)}, \sigma_\gamma^{2(r)})$, and
(a) Obtain $\beta^{(r+1)}$ by using the Monte Carlo expectation based Newton−Raphson iterative equation (see (4.28) and (4.50)):

$$\beta^{(r+1)} = \beta^{(r)} + E[\sum_{i=1}^{K} X_i' A_i(\gamma_i, \beta^{(r)}) X_i]^{-1} E[\sum_{i=1}^{K} X_i' \{y_i - \mu_i^*(\gamma_i, \beta^{(r)})\}|y], \tag{4.143}$$

where

$$\mu_i^*(\gamma_i, \beta^{(r)}) = [\mu_{i1}^*(\gamma_i, \beta^{(r)}), \ldots, \mu_{ij}^*(\gamma_i, \beta^{(r)}), \ldots, \mu_{in_i}^*(\gamma_i, \beta^{(r)})]'$$

and

$$A_i(\gamma_i, \beta^{(r)}) = \text{diag}[\mu_{i1}^*(\gamma_i, \beta^{(r)}), \ldots, \mu_{ij}^*(\gamma_i, \beta^{(r)}), \ldots, \mu_{in_i}^*(\gamma_i, \beta^{(r)})],$$

with $\mu_{ij}^*(\gamma_i, \beta) = \exp(x_{ij}'\beta + \gamma_i)$.
(b) Choose $\sigma_\gamma^{2(r+1)})$ to maximize $[1/N] \sum_{w=1}^{N} \log g(\gamma_i^{(w)}|\sigma_\gamma^2)$.
(c) Set $r = r + 1$.

3. If convergence is achieved, then declare $\beta^{(r+1)}$ and $\sigma_\gamma^{2(r+1)}$ to be the maximum likelihood estimates; otherwise, go back to Step 2.

Note that even though the semiparametric approach discussed in Section 4.4 and the Monte Carlo approach discussed in Section 4.5 are flexible on the distributional assumptions for the random effects, these techniques are, however, either long or numerically expensive. Furthermore, these approaches also have serious theoretical limitations, as they will not be applicable to clustered data such as longitudinal, where the correlations arise due to the dynamic relationship between observations at two different times (see Chapters 6 – 8), instead of common random effects shared by family members as in the present familial data setup. Consequently, these two approaches are further discussed only in Chapter 5 on familial models for binary data, but not in any other chapters dealing with longitudinal models.

Exercises

4.1. (Section 4.2.3) [Small σ_γ^2 based first four cumulants]
(a). For an auxiliary parameter t, derive $M_{\gamma_i}(t) = E[\exp(t\gamma_i)]$, the moment generating function (m.g.f.) of γ_i when its distribution is given by (4.33); that is,

$$g_w(\gamma_i) = \frac{\phi^\alpha}{\Gamma(\alpha)} \exp\{\alpha\gamma_i - \phi \exp(\gamma_i)\}.$$

(b). Show that the first four cumulants obtained from $\log M_{\gamma_i}(t)$, the cumulants generating function (c.g.f) of γ_i, are given by

$$K_1^* = \psi(\alpha) - \log\phi, \quad K_2^* = \psi'(\alpha), \quad K_3^* = \psi''(\alpha), \quad K_4^* = \psi'''(\alpha),$$

where $\psi(\alpha)$ is the digamma function

$$\psi(\alpha) = \frac{\partial\Gamma(\alpha)}{\partial\alpha} = -\xi - \frac{1}{\alpha} + \sum_{j=1}^{\infty} \frac{\alpha}{j(\alpha+j)}$$

with $\xi = 0.57721$, Euler's constant, and where $\psi'(\alpha)$, $\psi''(\alpha)$, and $\psi'''(\alpha)$ are, respectively, the first−, second−, and the third-order derivatives of $\psi(\alpha)$ with respect to α.
(c). Recall that α and ϕ in (4.33), that is, in the density $g_w(\gamma_i)$, are chosen such that

$$K_1^* = 0 \quad \text{and} \quad K_2^* = \sigma_\gamma^2.$$

Use this argument and justify that

$$K_3^* = -\sigma_\gamma^4 \quad \text{and} \quad K_4^* = 2\sigma_\gamma^6.$$

(d). Show that if γ_i has the normal distribution

$$g_N(\gamma_i) = (2\pi\sigma_\gamma^2)^{-1/2} \exp\{-\gamma_i^2/2\sigma_\gamma^2\},$$

then the first four cumulants of γ_i are given by

$$K_1 = 0, \quad K_2 = \sigma_\gamma^2, \quad K_3^* = 0, \quad K_4 = 0.$$

4.2. (Section 4.2.3) [Formulas for $P_\ell(\cdot)$ in (4.41)
Show that if $g_w(\gamma_i)$ given in Exercise 1 (a) [see also (4.33)] satisfies (4.41), then

$$\begin{aligned}
P_1(\gamma_i) &= \alpha - \phi\exp(\gamma_i)\\
P_2(\gamma_i) &= -\phi\exp(\gamma_i) + [P_1(\gamma_i)]^2\\
P_3(\gamma_i) &= -\phi\exp(\gamma_i) - 3\lambda\exp(\gamma_i)[P_1(\gamma_i)] + [P_1(\gamma_i)]^3\\
P_4(\gamma_i) &= -\phi\exp(\gamma_i) + 3\lambda^2\exp(2\gamma_i) - 4\phi\exp(\gamma_i)P_1(\gamma_i)\\
&\quad -6\phi\exp(\gamma_i)[P_1(\gamma_i)]^2 + [P_1(\gamma_i)]^4.
\end{aligned}$$

4.3. (Section 4.3.3.1) [Formulas for ξ_1, ξ_2, and ξ_3 in (4.95)]
Express the conditional Poisson mean μ^*_{ij} in (4.92) as $\mu^*_{ij} = m_{ij}a_ib_i$, where

$$m_{ij} = \exp(x'_{ij}\beta),\quad a_i = \exp(z_{i1}\sigma_\gamma\gamma^*_i),\ \text{ and }\ b_i = \exp(z_{i2}\sigma_\tau\tau^*_i).$$

For

$$\gamma^*_i \sim \mathrm{N}(0,1)\ \text{ and }\ \tau^*_i \sim \mathrm{N}(0,1),$$

and γ_i and τ_i are independent for all $i = 1,\ldots,K$, show that

$$\xi_1 = E[W_1] = \sum_{i=1}^K \tilde{\xi}_{i1},\ \ \xi_2 = E[W_2] = \sum_{i=1}^K z_{i1}\tilde{\xi}_{i2},;\ \text{and}\ \ \xi_3 = E[W_3] = \sum_{i=1}^K z_{i2}\tilde{\xi}_{i2},$$

where

$$\tilde{\xi}_{i1} = \sum_{j=1}^{n_i} x_{ij}\mu_{ij},\quad \tilde{\xi}_{i2} = \sum_{j=1}^{n_i}\lambda_{ijj} + 2\sum_{j<k}^{n_i}\lambda_{ijk},$$

with

$$\begin{aligned}
\mu_{ij} = E(Y_{ij}) &= m_{ij}p_iq_i\\
\lambda_{ijj} = E(Y^2_{ij}) &= m_{ij}p_iq_i(1 + m_{ij}p^3_iq^3_i)\\
\lambda_{ijk} = E(Y_{ij}Y_{ik}) &= m_{ij}m_{ik}p^4_iq^4_i,
\end{aligned}$$

where

$$p_i = \mathrm{E}(a_i) = \exp(z^2_{i1}\sigma^2_\gamma/2),\ \text{ and }\ q_i = \mathrm{E}(b_i) = \exp(z^2_{i2}\sigma^2_\tau/2).$$

4.4. (Section 4.3.3.1) [Formulas for the elements of the derivative matrix **P** in (4.97)]
For ξ_1, ξ_2, and ξ_3 given as in Exercise 4.3, verify that their derivatives:
with respect to β are−

$$\frac{\partial\xi_1}{\partial\beta'} = \sum_{i=1}^K\sum_{j=1}^{n_i} x_{ij}x'_{ij}\mu_{ij},\qquad \frac{\partial\xi_2}{\partial\beta'} = \sum_{i=1}^K z_{i1}\xi^*_{i1},\qquad \frac{\partial\xi_3}{\partial\beta'} = \sum_{i=1}^K z_{i2}\xi^*_{i1};$$

with respect to σ^2_γ are−

$$\frac{\partial \xi_1}{\partial \sigma_\gamma^2} = \frac{1}{2}\sum_{i=1}^{K} z_{i1}^2 \tilde{\xi}_{i1}, \qquad \frac{\partial \xi_2}{\partial \sigma_\gamma^2} = \sum_{i=1}^{K} z_{i1}^3 \xi_{i2}^*, \qquad \frac{\partial \xi_3}{\partial \sigma_\gamma^2} = \sum_{i=1}^{K} z_{i1}^2 z_{i2} \xi_{i2}^*;$$

with respect to σ_τ^2 have the formulas–

$$\frac{\partial \xi_1}{\partial \sigma_\tau^2} = \frac{1}{2}\sum_{i=1}^{K} z_{i2}^2 \tilde{\xi}_{i1}, \qquad \frac{\partial \xi_2}{\partial \sigma_\tau^2} = \sum_{i=1}^{K} z_{i1} z_{i2}^2 \xi_{i2}^*, \qquad \frac{\partial \xi_3}{\partial \sigma_\tau^2} = \sum_{i=1}^{K} z_{i2}^3 \xi_{i2}^*,$$

where $\tilde{\xi}_{i1}$ is as in Exercise 4.3, and ξ_{i1}^* and ξ_{i2}^* are given by

$$\begin{aligned}\xi_{i1}^* &= \sum_{j=1}^{n_i} x_{ij}' \mu_{ij}(1+2\mu_{ij} p_i^2 q_i^2) + 2\sum_{j<k}^{n_i} \lambda_{ijk}(x_{ij}' + x_{ik}') \\ \xi_{i2}^* &= \sum_{j=1}^{n_i} \frac{\mu_{ij}}{2}(1+4\mu_{ij} p_i^2 q_i^2) + 4\sum_{j<k}^{n_i} \lambda_{ijk}.\end{aligned}$$

4.5. (Section 4.3.3.2) [Formulas for the elements of the derivative matrix $\partial \zeta_i'/\partial \theta$: $(p+2) \times [n_i(n_i+3)/2]$ in (4.98)]
For $\theta = (\beta', \sigma_\gamma^2, \sigma_\tau^2)'$, and $\zeta_i = (\mu_i', \lambda_{i1}', \lambda_{i2}')'$ given as in (4.99), verify that their derivatives:
with respect to β are–

$$\frac{\partial \mu_{ij}}{\partial \beta} = \mathbf{x}_{ij}\mu_{ij}, \quad \frac{\partial \lambda_{ijj}}{\partial \beta} = \mathbf{x}_{ij}\mu_{ij}(1+2\mathbf{x}_{ij}\mu_{ij}p_i^2 q_i^2), \quad \frac{\partial \lambda_{ijk}}{\partial \beta} = (\mathbf{x}_{ij} + \mathbf{x}_{ik})\mu_{ij}\mu_{ik}p_i^2 q_i^2;$$

with respect to σ_γ^2 are–

$$\frac{\partial \mu_{ij}}{\partial \sigma_\gamma^2} = \frac{z_{i1}^2}{2}\mu_{ij}, \quad \frac{\partial \lambda_{ijj}}{\partial \sigma_\gamma^2} = \frac{z_{i1}^2}{2}\mu_{ij}(1+4\mu_{ij}p_i^2 q_i^2), \quad \frac{\partial \lambda_{ijk}}{\partial \sigma_\gamma^2} = 2z_{i1}^2 \mu_{ij}\mu_{ik}p_i^2 q_i^2;$$

with respect to σ_τ^2 have the formulas–

$$\frac{\partial \mu_{ij}}{\partial \sigma_\tau^2} = \frac{z_{i2}^2}{2}\mu_{ij}, \quad \frac{\partial \lambda_{ijj}}{\partial \sigma_\tau^2} = \frac{z_{i2}^2}{2}\mu_{ij}(1+4\lambda_{ij}p_i^2 q_i^2), \quad \frac{\partial \lambda_{ijk}}{\partial \sigma_\tau^2} = 2z_{i2}^2 \mu_{ij}\mu_{ik}p_i^2 q_i^2.$$

4.6. (Section 4.4) [Formulas for first eight moments of y_{ij} with p.d.f. (4.120)]
Let $m_{ij,1}' = E_{\exp}(y_{ij})$ and $m_{ij,s} = E_{\exp}(y_{ij} - m_{ij,1}')^s$ for $s = 2, \ldots, 8$, with pdf of y_{ij} as in (4.120). Show that the m.g.f. (see also (4.9) under Lemma 4.1] for Y_{ij} with p.d.f. (4.120) has the formula

$$M_{Y_{ij}|\theta_{ij}}(s) = E[\exp(sY_{ij})] = \exp[a(\theta_{ij}+s) - a(\theta_{ij}],$$

with $a(\theta_{ij}) = \exp(\theta_{ij})$, and show that

$$m_{ij,1}' = a_{ij}', \;\; m_{ij,2} = a_{ij}'', \;\; m_{ij,3} = a_{ij}''', \;\; m_{ij,4} = a_{ij}^{IV} + 3m_{ij,2}^2$$

$$m_{ij,5} = a_{ij}^{V} + 10m_{ij,2}m_{ij,3}, \;\; m_{ij,6} = a_{ij}^{IV} + 15m_{ij,2}m_{ij,4} + 10m_{ij,3}^{2} - 30m_{ij,2}^{3}$$

$$m_{ij,7} = a_{ij}^{VII} + 21m_{ij,2}m_{ij,5} + 35m_{ij,3}m_{ij,4} - 210m_{ij,2}^{2}m_{ij,3}$$

$$m_{ij,8} = a_{ij}^{VIII} + 28m_{ij,2}m_{ij,6} + 56m_{ij,3}m_{ij,5}$$

$$-630m_{ij,2}^{2}m_{ij,4} + 70m_{ij,4}^{2} - 560m_{ij,2}m_{ij,3}^{2} + 945m_{ij,2}^{4},$$

where a'_{ij}, $a''_{ij}, \ldots, a_{ij}^{VIII}$, respectively, denote the first−, second−, ..., eighth-order derivative of $a(\theta_{ij})$ in (4.120) with respect to θ_{ij}.

References

1. Breslow, N, E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. . *J. Amer. Stat. Assoc.,* 88, 9 – 25.
2. Breslow, N. E. & Lin, X. (1995). Bias correction in generalized linear models with a single component of dispersion. *Biometrika,* 82, 81 – 92.
3. Chowdhury, M. R. I. & Sutradhar, B. C. (2009). Generalized quasilikelihood versus hierarchial likelihood inferences in generalized linear mixed models for count data. *Sankhya B: Indian J. Statist.* 71, 1 – 15.
4. Fahmeir, L. & Tutz, G. T. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* New York: Springer.
5. Gelfand, A. E. & Carlin, B. P. (1993). Maximum likelihood estimation for constrained or missing-data problems. *Canad. J. Statist.,* 21, 303 – 311.
6. Geyer, C. J. & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Stats. Soc. B,* 54, 657 – 699.
7. Henderson, C. R. (1950). Estimation of genetic parameters (Abstract). *Ann. Math. Stat.,* 21, 309 – 310.
8. Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.,* 93, 720 – 729.
9. Jiang, J. & Zhang, W. (2001). Robust estimation in generalized linear mixed models. *Biometrika,* 88, 753 – 765.
10. Johnson, N. L. & Kotz, S. (1972). *Continuous Univariate Distributions* 1. New York: Wiley.
11. Jowaheer, V., Sutradhar, B. C., & Sneddon, G. (2009). On familial Poisson mixed models with multi-dimensional random effects. *J. Statist. Comp. Simul.,* 79, 1043 – 1062.
12. Lee, Y. & Nelder, J. (1996). Hierarchical generalized linear models. *J. R. Stats. Soc. B,* 58, 619 – 678.
13. Lin, X. (1997). Variance component testing in generalized linear models with random effects. *Biometrika,* 84, 309 – 326.
14. McCullagh, P. (1983). Quasilikelihood functions. *Ann. Statist.,* 11, 59 – 67.
15. McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models,* 2nd edn. London: Chapman and Hall.
16. McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *J. Amer. Statist. Assoc.,* 92, 162 – 170.
17. Nelder, J. & Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. A,* 135, 370 – 384.
18. Patterson, H. D. & Thompson, R. (1974). Recovery of interblock information when block sizes are unequal. *Biometrika,* 58, 545 – 554.

19. Searle, S. E., Casella, G. & McCulloch, C. E. (1992). *Variance Components.* New York: John Wiley & Sons.
20. Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statist. Sci.,* 18, 377 – 393.
21. Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhya B: Indian J. Stat.,* 66, 261 – 289.
22. Sutradhar, B. C. & Das, K. (2001). On higher order approximation to the likelihood inference in the Poisson mixed model. *Statist. Prob. Lett.,* 52, 59 – 67.
23. Sutradhar, B. C. & Qu, Z. (1998). On approximate likelihood inference in Poisson mixed model. *Canad. J. Ststist.,* 26, 169 – 186.
24. Sutradhar, B. C. & Rao, R. Prabhakar (2001). On marginal quasi-likelihood inference in generalized linear mixed models. *J. Multivariate Anal.,* 76, 1 – 34.
25. Sutradhar, B. C. & Rao, R. Prabhakar (2003). On quasilikelihood inference in generalized linear mixed models with two variance components. *Canad. J. Statist.,* 31, 415 – 435.
26. Tanner, M. A. (1993). *Tools for Statistical Inference: Observed Data and Data Augmentation.* Berlin: Springer-Verlag.
27. Ten Have, T. R. & Morabia, A. (1999). Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometrics,* 55, 85 – 93.
28. Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalised linear models, and the Gauss-Newton method. *Biometrika,* 61, 439 – 447.

Chapter 5
Familial Models for Binary Data

As opposed to Chapter 4, we now consider y_{ij} as the binary response for the jth $(j = 1, \ldots, n_i)$ member of the ith $(i = 1, \ldots, K)$ family/cluster. Suppose that $x_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ is the p-dimensional covariate vector associated with the binary response y_{ij}. For example, in a chronic obstructive pulmonary disease (COPD) study, y_{ij} denotes the impaired pulmonary function (IPF) status (yes or no), and x_{ij} is the vector of covariates such as gender, race, age, and smoking status, for the jth sibling of the ith COPD patient. Note that in this problem it is likely that the IPF status for n_i siblings of the ith patient may be influenced by an unobservable random effect (γ_i) due to the ith COPD patient. Similar to the Poisson mixed model discussed in the last chapter, let γ_i denote this random effect. This common random effect makes the binary responses of any two siblings of the same patient correlated, and this correlation is referred to as the familial correlation. It is of scientific interest to find the effects of the covariates on the binary responses, that is, IPF status of an individual sibling after taking the familial correlations (caused by the variation in random effects) into account.

In Section 5.1, we provide the marginal (unconditional) distributional properties of the binary response variable y_{ij}, and the unconditional familial correlation structure for the responses of the members of the ith familiy under the assumption that the random family effects follow a normal distribution. Note that unlike the Poisson mixed model case, one cannot obtain any explicit formulas for the moments of the binary variable in the mixed model setup even if the random effects are assumed to be normally distributed. Thus, the basic properties discussed below in Section 5.1 are developed based on a suitable numerical such as simulation or binomial approximation approach. In Section 5.2 we discuss various inference techniques that produce at least consistent estimates for the parameters involved. The penalized quasi-likelihood (PQL) [Breslow and Clayton (1993)] and hierarchical likelihood (HL) [Lee and Nelder (1996)] approaches were found to have inconsistency problems for the inferences in the Poisson case, and because there is no reason why they will do better in the binary case, we therefore do not include these approaches in Section 5.2. To be specific, we provide details on the development of estimating equations using the method of moments (MM) [Jiang (1998), Jiang and Zhang

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*, Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_5,

(2001)], generalized quasi-likelihood (GQL) [Sutradhar (2004)], and the exact maximum likelihood (ML) [Sutradhar and Mukerjee (2005)] approaches. This is done for the cases when the mixed model contains random effects from a single source. In Section 5.3, we provide the GQL inferences for the binary mixed models with normal random effects but from two sources [Sutradhar and Rao (2003)].

In Section 5.4, we highlight a semiparametric (SP) estimation approach, whereas in Section 5.5 we discuss a Monte Carlo (MC) based likelihood approximation. Note that these two approaches correspond to those for the count data case discussed in Sections 4.4 and 4.5, respectively. These two approaches do not require the assumption that the random effects follow a Gaussian distribution. However, they are computationally expensive, and also they are not directly useful for inferences in the dynamic models based longitudinal data analysis.

5.1 Binary Mixed Models and Basic Properties

Let $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$ be the $n_i \times 1$ vector of binary responses from n_i members of the ith $(i = 1, \ldots, K)$ family. Let β be a $p \times 1$ vector of unknown fixed effects of x_{ij} on y_{ij}, x_{ij} being the p-dimensional covariate vector for the jth $(j = 1, \ldots, n_i)$ member of the ith family. Suppose that conditional on the random family effect γ_i, n_i counts due to the ith family are independent. The data of this type can be modelled as

$$f(y_i|\gamma_i) = \Pi_{j=1}^{n_i} \left[\{\pi_{ij}^*\}^{y_{ij}} \{1 - \pi_{ij}^*\}^{1-y_{ij}} \right], \tag{5.1}$$

where

$$\pi_{ij}^* = Pr[Y_{ij} = 1|\gamma_i] = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}$$

with $\eta_{ij}(\beta, \gamma_i) = x_{ij}'\beta + \gamma_i$. Note that, similar to (4.1), this conditional joint density may be written as

$$\begin{aligned} f(y_i|\gamma_i) &= \exp\left[\sum_{j=1}^{n_i} y_{ij}\eta_{ij} - \sum_{j=1}^{n_i} a(\eta_{ij}) \right] \\ &= \exp\left[\sum_{j=1}^{n_i} y_{ij}\eta_{ij} - \sum_{j=1}^{n_i} \log\{1 + \exp(\eta_{ij})\} \right]. \end{aligned} \tag{5.2}$$

Further note that if the random effect γ_i is assumed to have an unspecified distribution with mean 0 and variance σ_γ^2 and γ_is are independent, that is, $\gamma_i \stackrel{\text{iid}}{\sim} (0, \sigma_\gamma^2)$, then for $\gamma_i^* = \gamma_i/\sigma_\gamma$, the linear predictor η_{ij} in (5.1) and (5.2) may be expressed as

$$\eta_{ij}(\beta, \sigma_\gamma, \gamma_i^*) = x_{ij}'\beta + \gamma_i = x_{ij}'\beta + \sigma_\gamma\gamma_i^*, \tag{5.3}$$

where $\gamma_i^* \stackrel{\text{iid}}{\sim} (0, 1)$.

Before proceeding toward the development of estimation techniques for the parameters β and σ_γ^2, involved in the binary mixed model (5.1) – (5.3), we provide below the basic properties such as the unconditional marginal and product moments of orders two, three and four. The marginal and product moments of order two are helpful in understanding the mean and correlation structures of the model, and also these moments along with the product moments of order three and four are exploited to develop the desired MM and GQL estimation approaches.

Lemma 5.1. Conditional on γ_i^*, the mean and the variance of Y_{ij}, and the covariances between Y_{ij} and Y_{ik} for $j \neq k,\ j,\ k = 1,\ldots,n_i$ are given by

$$E[Y_{ij}|\gamma_i^*] = \pi_{ij}^*(\gamma_i^*) = \frac{\exp(x_{ij}'\beta + \sigma_\gamma\gamma_i^*)}{1+\exp(x_{ij}'\beta + \sigma_\gamma\gamma_i^*)} \tag{5.4}$$

$$\text{var}[Y_{ij}|\gamma_i^*] = \pi_{ij}^*(\gamma_i^*)\{1-\pi_{ij}^*(\gamma_i^*)\} \tag{5.5}$$

$$\text{cov}[(Y_{ij},Y_{i,k})|\gamma_i^*] = 0, \tag{5.6}$$

and, for $\gamma_i^* \overset{\text{iid}}{\sim} N(0,1)$, the corresponding unconditional mean, variance, and the covariances are given by

$$E[Y_{ij}] = \pi_{ij}(\beta,\sigma_\gamma^2) = \int \pi_{ij}^*(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* \tag{5.7}$$

$$\text{var}[Y_{ij}] = \sigma_{ijj}(\beta,\sigma_\gamma^2) = \pi_{ij}(\beta,\sigma_\gamma^2)(1-\pi_{ij}(\beta,\sigma_\gamma^2)) \tag{5.8}$$

$$\begin{aligned}\text{cov}[Y_{ij},Y_{ik}] = \sigma_{ijk}(\beta,\sigma_\gamma^2) &= \int \pi_{ij}^*(\gamma_i^*)\pi_{ik}^*(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* - \pi_{ij}\pi_{ik} \\ &= \lambda_{ijk} \text{ (say)} - \pi_{ij}\pi_{ik},\end{aligned} \tag{5.9}$$

with $g_N(\gamma_i^*|1)$ as the standard normal density, yielding the pairwise familial correlations as

$$\text{corr}[Y_{ij},Y_{ik}] = \frac{\sigma_{ijk}(\beta,\sigma_\gamma^2)}{[\pi_{ij}(\beta,\sigma_\gamma^2)(1-\pi_{ij}(\beta,\sigma_\gamma^2))\pi_{ik}(\beta,\sigma_\gamma^2)(1-\pi_{ik}(\beta,\sigma_\gamma^2))]^{1/2}}. \tag{5.10}$$

Proof: Because $f(y_{ij}|\gamma_i^*) = \{\pi_{ij}^*\}^{y_{ij}}\{1-\pi_{ij}^*\}^{1-y_{ij}}$ by (5.1), it then follows that

$$E[Y_{ij}|\gamma_i^*] = \sum_{y_{ij}=0}^{1} y_{ij}\{\pi_{ij}^*\}^{y_{ij}}\{1-\pi_{ij}^*\}^{1-y_{ij}},$$

yielding the conditional mean as in (5.4). Note that for the binary data, for any finite integer r, it also follows that

$$E[Y_{ij}^r|\gamma_i^*] = E[Y_{ij}|\gamma_i^*] = \pi_{ij}^*(\gamma_i^*), \tag{5.11}$$

yielding the conditional variance as in (5.5). Furthermore, because conditional on γ_i^*, Y_{ij} and Y_{ik} are independent, it then follows that

$$E[Y_{ij}Y_{ik}|\gamma_i^*] = E[Y_{ij}|\gamma_i^*]E[Y_{ik}|\gamma_i^*] = \pi_{ij}^*(\gamma_i^*)\pi_{ik}^*(\gamma_i^*), \tag{5.12}$$

yielding the conditional covariance as in (5.6).

Next, one obtains the unconditional mean, variance, and the covariance in (5.7) – (5.9) by using

$$E[Y_{ij}] = E_{\gamma_i^*}E[Y_{ij}|\gamma_i^*] = E_{\gamma_i^*}[\pi_{ij}^*(\gamma_i^*)] = \pi_{ij}(\beta, \sigma_\gamma^2) \tag{5.13}$$

$$E[Y_{ij}^2] = E_{\gamma_i^*}E[Y_{ij}^2|\gamma_i^*] = E_{\gamma_i^*}[\pi_{ij}^*(\gamma_i^*)] = \pi_{ij}(\beta, \sigma_\gamma^2) \tag{5.14}$$

$$E[Y_{ij}Y_{ik}] = E_{\gamma_i^*}E[Y_{ij}Y_{ik}|\gamma_i^*] = E_{\gamma_i^*}[\pi_{ij}^*(\gamma_i^*)\pi_{ik}^*(\gamma_i^*)] = \lambda_{ijk}(\beta, \sigma_\gamma^2). \tag{5.15}$$

In the manner similar to that for λ_{ijk}, the product moment of second order, one may compute the product moments of orders three and four for the binary data, as in the following lemma.

Lemma 5.2. Under the binary model (5.1) – (5.3), the unconditional product moments of orders three and four, are given by

$$E[Y_{ij}Y_{ik}Y_{i\ell}] = \delta_{ijk\ell} \text{ (say)} = \int \pi_{ij}^*(\gamma_i^*)\pi_{ik}^*(\gamma_i^*)\pi_{i\ell}^*(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^* \tag{5.16}$$

$$E[Y_{ij}Y_{ik}Y_{i\ell}Y_{im}] = \phi_{ijk\ell m} \text{ (say)} = \int \pi_{ij}^*(\gamma_i^*)\pi_{ik}^*(\gamma_i^*)\pi_{i\ell}^*(\gamma_i^*)\pi_{im}^*(\gamma_i^*) \times g_N(\gamma_i^*|1)d\gamma_i^*. \tag{5.17}$$

Proof: These third− and fourth-order product moments follow from the fact that conditional on γ_i, the binary responses from the members are independent. Thus,

$$\begin{aligned} E[Y_{ij}Y_{ik}Y_{i\ell}] &= E_{\gamma_i^*}E[Y_{ij}Y_{ik}Y_{i\ell}|\gamma_i^*] \\ &= E_{\gamma_i^*}[E(Y_{ij}|\gamma_i^*)E(Y_{ik}|\gamma_i^*)E(Y_{i\ell}|\gamma_i^*)] \\ &= E_{\gamma_i^*}[\pi_{ij}(\gamma_i^*)\pi_{ik}(\gamma_i^*)\pi_{i\ell}(\gamma_i^*)] \\ &= \int \pi_{ij}(\gamma_i^*)\pi_{ik}(\gamma_i^*)\pi_{i\ell}(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^*, \end{aligned} \tag{5.18}$$

where

$$\pi_{ij}(\gamma_i^*) = \frac{\exp[x_{ij}'\beta + \sigma_\gamma\gamma_i^*]}{1 + \exp[x_{ij}'\beta + \sigma_\gamma\gamma_i^*]}.$$

Similarly,

$$\begin{aligned} E[Y_{ij}Y_{ik}Y_{i\ell}Y_{im}] &= E_{\gamma_i^*}E[Y_{ij}Y_{ik}Y_{i\ell}Y_{im}|\gamma_i^*] \\ &= E_{\gamma_i^*}[E(Y_{ij}|\gamma_i^*)E(Y_{ik}|\gamma_i^*)E(Y_{i\ell}|\gamma_i^*)E(Y_{im}|\gamma_i^*)] \\ &= E_{\gamma_i^*}[\pi_{ij}(\gamma_i^*)\pi_{ik}(\gamma_i^*)\pi_{i\ell}(\gamma_i^*)\pi_{im}(\gamma_i^*)] \\ &= \int \pi_{ij}(\gamma_i^*)\pi_{ik}(\gamma_i^*)\pi_{i\ell}(\gamma_i^*)\pi_{im}(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^*. \end{aligned} \tag{5.19}$$

5.1.1 Computational Formulas for Binary Moments

Note that unlike the Poisson mixed model (see Lemmas 4.1 and 4.3), one can not obtain explicit formulas for the moments under the binary mixed models because of the integration difficulty of a complex function over the normal distribution for the random effects. Recall from Section 4.2.1 that a similar difficulty arose for the likelihood computation under the Poisson mixed models, where as a remedy the integrations were evaluated either by a simulation approach [Jiang (1998)] or by a binomial approximation approach [see Ten Have and Morabia (1999, eqn. (7)), for example]. We may use one of these approaches for the computation of the binary moments given in Lemmas 5.1 and 5.2.

Simulated Binary Moments

In the simulation technique, for a large N such as $N = 1000$, the first-order binary moment $\pi_{ij}(\beta, \sigma_\gamma^2)$ in (5.7) may be computed as

$$\pi_{ij}^{(s)}(\beta, \sigma_\gamma^2) = \frac{1}{N} \sum_{w=1}^{N} [\pi_{ij}^*(\gamma_{iw}^*)], \tag{5.20}$$

where γ_{iw}^* is a sequence of standard normal values for $w = 1, \ldots, N$. Next, by using the notations

$$\begin{aligned} \lambda_{ijk}^*(\gamma_i^*) &= \pi_{ij}^*(\gamma_i^*)\pi_{ik}^*(\gamma_i^*), \quad \delta_{ijk\ell}^*(\gamma_i^*) = \pi_{ij}^*(\gamma_i^*)\pi_{ik}^*(\gamma_i^*)\pi_{i\ell}^*(\gamma_i^*), \\ \phi_{ijk\ell m}^*(\gamma_i^*) &= \pi_{ij}^*(\gamma_i^*)\pi_{ik}^*(\gamma_i^*)\pi_{i\ell}^*(\gamma_i^*)\pi_{im}^*(\gamma_i^*), \end{aligned}$$

one may follow (5.9), (5.16) and (5.17), and compute the simulation based binary product moments of orders two, three, and four, as

$$\lambda_{ijk}^{(s)}(\beta, \sigma_\gamma^2) = E[Y_{ij}Y_{ik}] = \frac{1}{N} \sum_{w=1}^{N} \lambda_{ijk}^*(\gamma_{iw}^*) \tag{5.21}$$

$$\delta_{ijk\ell}^{(s)}(\beta, \sigma_\gamma^2) = E[Y_{ij}Y_{ik}Y_{i\ell}] = \frac{1}{N} \sum_{w=1}^{N} \delta_{ijk\ell}^*(\gamma_{iw}^*) \tag{5.22}$$

$$\phi_{ijk\ell m}^{(s)}(\beta, \sigma_\gamma^2) = E[Y_{ij}Y_{ik}Y_{i\ell}Y_{im}] = \frac{1}{N} \sum_{w=1}^{N} \phi_{ijk\ell m}^*(\gamma_{iw}^*), \tag{5.23}$$

respectively.

Binary Moments Using Binomial Approximation to the Normal Integral

As an alternative to the simulation approach, one may compute the normal integrals in (5.7), (5.9), (5.16), and (5.17), by using the so-called binomial approximation. For a known reasonably big V such as $V = 5$, let $v_i \sim \text{binomial}(V, 1/2)$. Because γ_i^* has the standard normal distribution, consider

$$\gamma_i^* = \frac{v_i - V(1/2)}{V(1/2)(1/2)}.$$

One may then approximate the desired normal integral by a binomial approximation and compute the marginal and product moments up to order four as

$$\pi_{ij}^{(b)}(\beta, \sigma_\gamma^2) = \sum_{v_i=0}^{V} \pi_{ij}^*(v_i)] \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i} \tag{5.24}$$

$$\lambda_{ijk}^{(b)}(\beta, \sigma_\gamma^2) = \sum_{v_i=0}^{V} \lambda_{ijk}^*(v_i)] \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i} \tag{5.25}$$

$$\delta_{ijk\ell}^{(b)}(\beta, \sigma_\gamma^2) = \sum_{v_i=0}^{V} \delta_{ijk\ell}^*(v_i)] \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i} \tag{5.26}$$

$$\phi_{ijk\ell m}^{(b)}(\beta, \sigma_\gamma^2) = \sum_{v_i=0}^{V} \phi_{ijk\ell m}^*(v_i)] \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i}, \tag{5.27}$$

where

$$\pi_{ij}^*(v_i)] = \left[\pi_{ij}^*(\gamma_i^*)\}\right]_{[\gamma_i^* = \frac{v_i - V(1/2)}{V(1/2)(1/2)}]},$$

for example.

Note that these moments computed either based on the simulation approach or binomial approximation are used in the following sections for the inferences in the binary mixed model (5.1) – (5.3). As far as the estimation approach is concerned, as mentioned earlier, we concentrate on the MM, IMM, GQL, and ML approaches as they all produce consistent estimators for the parameters, some such as the GQL and ML being more efficient.

5.2 Estimation for Single Random Effect Based Parametric Mixed Models

5.2.1 Method of Moments (MM)

In this approach, similar to the Poisson mixed model [see eqns. (4.55) – (4.58)], one estimates β and σ_γ^2 by solving the moment equations

$$\psi_1(\beta, \sigma_\gamma^2) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij} \left\{ y_{ij} - \pi_{ij}^{(s)}(\beta, \sigma_\gamma^2) \right\} = 0, \tag{5.28}$$

and

$$\psi_2(\beta, \sigma_\gamma^2) = \sum_{i=1}^{K} \left[\sum_{j<k}^{n_i} y_{ij} y_{ik} - 2 \sum_{j<k}^{n_i} \lambda_{ijk}^{(s)}(\beta, \sigma_\gamma^2) \right] = 0 \tag{5.29}$$

[Jiang (1998)] respectively, where $\pi_{ij}^{(s)}(\beta,\sigma_\gamma^2)$ and $\lambda_{ijk}^{(s)}(\beta,\sigma_\gamma^2)$ are given by (5.20) and (5.21), respectively. Note that one could alternatively use $\pi_{ij}^{(b)}(\beta,\sigma_\gamma^2)$ (5.24) and $\lambda_{ijk}^{(b)}(\beta,\sigma_\gamma^2)$ (5.25), instead of $\pi_{ij}^{(s)}(\beta,\sigma_\gamma^2)$ and $\lambda_{ijk}^{(s)}(\beta,\sigma_\gamma^2)$, respectively, to construct the above two moment equations. Further note that because the first-order response y_{ij} and the squared response y_{ij}^2 in the binary case provide the same information, the joint estimating equations (5.28) and (5.29) under the binary mixed model are constructed by exploiting the first-order responses $y_{ij}(j=1,\ldots,n_i)$ and the product responses $y_{ij}y_{ik}(j\neq k,\ j,k=1,\ldots,n_i)$, whereas in the count data case all first-order $y_{ij}(j=1,\ldots,n_i)$, squared $y_{ij}^2(j=1,\ldots,n_i)$, and pairwise product $y_{ij}y_{ik}(j\neq k,\ j,k=1,\ldots,n_i)$ responses were exploited to form the estimating equations (4.55) and (4.56). This structural difference between the estimating equations for the count and binary data are also reflected under other methods such as IMM and GQL discussed in Sections 5.2.2 and 5.2.3, respectively.

For $\theta=[\beta',\sigma_\gamma]'$, one obtains the MM estimate by using the Gauss−Newton iterative equation

$$\hat{\theta}_{MM}(r+1)=\hat{\theta}_{MM}(r)+\left[\frac{\partial\xi'}{\partial\theta}\right]_r^{-1}[w-\xi]_r, \tag{5.30}$$

where $w=[w_1',w_2]'$, and $\xi=[\xi_1',\xi_2]'$, with

$$w_1=\sum_{i=1}^K\sum_{j=1}^{n_i}x_{ij}y_{ij},\quad w_2=\sum_{i=1}^K\sum_{j<k}^{n_i}y_{ij}y_{ik},$$

and

$$\xi_1=\sum_{i=1}^K\sum_{j=1}^{n_i}x_{ij}\pi_{ij}^{(s)}(\beta,\sigma_\gamma^2),\quad \xi_2=2\sum_{i=1}^K\left[\sum_{j<k}^{n_i}\lambda_{ijk}^{(s)}(\beta,\sigma_\gamma^2)\right], \tag{5.31}$$

and where $[]_r$ denotes that the expression within the square bracket is evaluated at $\theta=\hat{\theta}_{MM}(r)$, the estimate obtained for the rth iteration. Let the final solution obtained from (5.30) be denoted by $\hat{\theta}_{MM}$.

Note that under the binary mixed model, the computation for the derivative $\partial\xi'/\partial\theta$ is slightly more complicated than the Poisson mixed model case. For convenience, by (5.4) and (5.21), we provide the formulas for the associated derivatives as follows.

$$\begin{aligned}\frac{\partial\pi_{ij}^{(s)}(\beta,\sigma_\gamma^2)}{\partial\beta}&=\frac{1}{N}\sum_{w=1}^N\frac{\partial\pi_{ij}^*(\gamma_{iw}^*)}{\partial\beta}\\&=\frac{1}{N}\sum_{w=1}^N\pi_{ij}^*(\gamma_{iw}^*)[1-\pi_{ij}^*(\gamma_{iw}^*)]x_{ij}\end{aligned} \tag{5.32}$$

$$\frac{\partial\pi_{ij}^{(s)}(\beta,\sigma_\gamma^2)}{\partial\sigma_\gamma}=\frac{1}{N}\sum_{w=1}^N\frac{\partial\pi_{ij}^*(\gamma_{iw}^*)}{\partial\beta}$$

$$= \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw})[1-\pi^*_{ij}(\gamma^*_{iw})]\gamma^*_{iw}, \tag{5.33}$$

and

$$\begin{aligned}
\frac{\partial \lambda^{(s)}_{ijk}(\beta,\sigma^2_\gamma)}{\partial \beta} &= \frac{1}{N}\sum_{w=1}^{N}\frac{\partial \lambda^*_{ijk}(\gamma^*_{iw})}{\partial \beta}\\
&= \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw})\pi^*_{ik}(\gamma^*_{iw})[2-\pi^*_{ij}(\gamma^*_{iw})-\pi^*_{ik}(\gamma^*_{iw})]x_{ij} \qquad (5.34)\\
\frac{\partial \lambda^{(s)}_{ijk}(\beta,\sigma^2_\gamma)}{\partial \sigma_\gamma} &= \frac{1}{N}\sum_{w=1}^{N}\frac{\partial \lambda^*_{ijk}(\gamma^*_{iw})}{\partial \sigma_\gamma}\\
&= \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw})\pi^*_{ik}(\gamma^*_{iw})[2-\pi^*_{ij}(\gamma^*_{iw})-\pi^*_{ijk}(\gamma^*_{iw})]\gamma^*_{iw}. \qquad (5.35)
\end{aligned}$$

This completes the construction of the iterative equation (5.30).

Further note that becuase $E[W-\xi]=0$, the MM estimator $\hat{\theta}_{MM}$ obtained from (5.30) is consistent for θ but it may still produce biased estimators in finite sample cases. Moreover, the MM estimator can be inefficient. As far as the asymptotic variance of $\hat{\theta}_{MM}$ is concerned, one may obtain this from the fact that as $K\to\infty$, it follows from the multivariate central limit theorem [Mardia, Kent and Bibby (1979, p. 51), for example] that $\hat{\theta}_{MM}$ has the multivariate Gaussian distribution with mean θ and the variance given by

$$\mathrm{var}(\hat{\theta}_{MM}) = \mathrm{limit}_{K\to\infty}\left[\frac{\partial \xi'}{\partial \theta}\right]^{-1} V \left[\frac{\partial \xi}{\partial \theta'}\right]^{-1}, \tag{5.36}$$

where $V=\mathrm{var}[W-\xi]=\mathrm{var}(W)$.

5.2.2 *An Improved Method of Moments (IMM)*

Note that as the moment equations in (5.28) – (5.29) do not exploit the covariances (5.9) or correlations (5.10) among the members of the family, they produce inefficient estimates. As an improvement over this approach, Jiang and Zhang (2001), for example, have used an improved method of moments (IMM) estimation that solves the moment estimating equation

$$B[w^*-\psi^*]=0, \tag{5.37}$$

instead of $[w-\xi]=0$ in (5.28) – (5.29), where

$$w^* = [w'_{11},\, w_{21},\ldots,w_{2i},\ldots,w_{2K}]'$$

$$= [\sum_{i=1}^{K}\sum_{j=1}^{n_i} x'_{ij}y_{ij}, \sum_{j\neq k}^{n_1} y_{1j}y_{1k}, \ldots, \sum_{j\neq k}^{n_i} y_{ij}y_{ik}, \ldots, \sum_{j\neq k}^{n_K} y_{Kj}y_{Kk}]', \tag{5.38}$$

and

$$\begin{aligned}\psi^* &= [E(W'_{11}), E(W_{21}), \ldots, E(W_{2i}), \ldots, E(W_{2K})]' \\ &= [\sum_{i=1}^{K}\sum_{j=1}^{n_i} x'_{ij}\pi_{ij}(\beta,\sigma_\gamma^2), \sum_{j\neq k}^{n_1} \lambda_{1jk}(\beta,\sigma_\gamma^2), \ldots, \sum_{j\neq k}^{n_i} \lambda_{ijk}(\beta,\sigma_\gamma^2), \\ &\quad \ldots, \sum_{j\neq k}^{n_K} \lambda_{Kjk}(\beta,\sigma_\gamma^2)]',\end{aligned} \tag{5.39}$$

with $\pi_{ij}(\beta,\sigma_\gamma^2)$ and $\lambda_{ijk}(\beta,\sigma_\gamma^2)$ as in (5.13) and (5.15), respectively. Also, in (5.37), $B = D'V^{*-1}$ with D as the derivative matrix $D = \partial\psi^*/\partial\theta$, and V^* is the covariance matrix of W^*. Note that as far as the dimension is concerned, D in (5.37) is the $(p+K)\times(p+1)$ derivative matrix and V^* is the $(p+K)\times(p+K)$ covariance matrix of W^*.

In principle, the construction of the V^* matrix must require the computations for the third− and fourth-order moments of the responses in the same cluster. This is because, for the computations of $\text{var}(W_{2i})(i = 1,\ldots,K)$, one requires to compute $\text{cov}(Y_{i1}Y_{i2}, Y_{i2}Y_{i3})$ and $\text{cov}(Y_{i1}Y_{i2}, Y_{i3}Y_{i4})$, for example. Here

$$\text{cov}(Y_{i1}Y_{i2}, Y_{i2}Y_{i3}) = E(Y_{i1}Y_{i2}Y_{i3}) - E(Y_{i1}Y_{i2})E(Y_{i2}Y_{i3}),$$

and

$$\text{cov}(Y_{i1}Y_{i2}, Y_{i3}Y_{i4}) = E(Y_{i1}Y_{i2}Y_{i3}Y_{i4}) - E(Y_{i1}Y_{i2})E(Y_{i3}Y_{i4}),$$

which clearly require the computations of the third-order moment $E(Y_{i1}Y_{i2}Y_{i3})$ and the fourth-order moment $E(Y_{i1}Y_{i2}Y_{i3}Y_{i4})$, respectively. But, as the computation of the fourth-order moments matrix V^*, and hence the computation of the so-called optimal $B = D'V^{*-1}$ matrix is complicated, Jiang and Zhang (2001, Section 3, p. 758) suggest using a simple form for the B matrix, say B_0, which is free from higher-order moments such as moments of orders three and four. Thus, instead of (5.37), they suggest solving an estimating equation

$$B_0(w^* - \psi^*) = 0, \tag{5.40}$$

where B_0 is an alternative choice for the so-called optimal B matrix in (5.37), which is free from higher-order moments. It has been, however, demonstrated by Sutradhar (2004) that there cannot be any such matrix free from moments of orders three and four. We provide this interesting contradiction below.

5.2.2.1 Can There Be an Optimal B Free from Third− and Fourth-Order Moments Under Simple Binary Logistic Mixed Models?

We examine whether there can be any choice for the B matrix in (5.37) which is optimal but does not involve any moments higher than second order. For simplicity, this we do, in connection with the simple binary logistic model considered by Jiang and Zhang (2001, pp. 756 – 57) under a balanced mixed model with $n_i = n$, n_i being the size of the ith $(i = 1, \ldots, K)$ cluster. Next, in the following section, we show the effect of the mis-specification of the B matrix in estimating the parameter vector θ. This we do by obtaining the estimate of θ from (5.40) when, in fact, (5.37) with $B = D'V^{*-1}$, is the true improved moment estimating equation for θ.

Let y_{ij} be the binary responses of the jth $(j = 1, \ldots, n)$ individual of the ith $(i = 1, \ldots, K)$ cluster. Similar to Jiang and Zhang (2001, pp. 756 – 57), consider a simple binary logistic model with $x_{ij} = x_i$ for all $j = 1, \ldots, n$. Also, consider, $p = 1$. It then follows from (5.1) – (5.3) that

$$\text{logit}\{\Pr(y_{ij} = 1|\gamma_i^*)\} = x_i\beta + \sigma_\gamma\gamma_i^*.$$

Here, $\gamma_i^* \stackrel{iid}{\sim} N(0,1)$. To estimate $\theta = (\beta, \sigma_\gamma)'$, Jiang and Zhang (2001, §2.3) have chosen

$$\begin{aligned} B_0 = & \ \text{diag}(I_1, 1'_K) \\ = & \begin{bmatrix} 1 \ 0 \ \ldots \ 0 \\ \\ 0 \ 1 \ \ldots \ 1 \end{bmatrix}, \end{aligned} \tag{5.41}$$

as an optimal choice for the B matrix. We now verify whether this special B_0 matrix, not only free from higher-order moments but free from any moments, is really equal to the optimal $B = D'V^{*-1}$ matrix under the present simple binary logistic model.

Because $x_{ij} = x_i$, using a simpler notation $h_i(x_i\beta + \sigma_\gamma\gamma_i^*) = \pi_{ij}^*(\gamma_i^*) = \exp(x_i\beta + \sigma_\gamma\gamma_i^*)/\{1 + \exp(x_i\beta + \sigma_\gamma\gamma_i^*)\}$, it follows from (5.13), (5.15), and (5.39) that the derivative of the ψ^* vector with respect to $\theta = (\beta, \sigma)'$ is given by

$$\begin{aligned} D' &= \begin{bmatrix} \frac{\partial E(W_{11})}{\partial \beta} & \frac{\partial E(W_{21})}{\partial \beta} & \cdots & \frac{\partial E(W_{2K})}{\partial \beta} \\ \\ \frac{\partial E(W_{11})}{\partial \sigma} & \frac{\partial E(W_{21})}{\partial \sigma} & \cdots & \frac{\partial E(W_{2K})}{\partial \sigma} \end{bmatrix} \\ &= n \begin{bmatrix} \sum_{i=1}^K x_i e_{1i}^* & 2(n-1)e_{31}^* & \ldots \ 2(n-1)e_{3K}^* \\ \\ \sum_{i=1}^K x_i e_{2i}^* & 2(n-1)e^*41 & \ldots \ 2(n-1)e_{4K}^* \end{bmatrix}, \end{aligned} \tag{5.42}$$

where e_{1i}^*, e_{2i}^*, e_{3i}^*, and e_{4i}^* are defined as

$$e_{1i}^* = E_{\gamma_i^*}[h_i(\cdot)\{1 - h_i(\cdot)\}x_i], \quad e_{2i}^* = E_{\gamma_i^*}[\gamma_i^* h_i(\cdot)\{1 - h_i(\cdot)\}],$$

$$e_{3i}^* = E_{\gamma_i^*}[h_i^2(\cdot)\{1 - h_i(\cdot)\}x_i], \ \text{ and } \ e_{4i}^* = E_{\gamma_i^*}[\gamma_i^* h_i^2(\cdot)\{1 - h_i(\cdot)\}],$$

respectively.

Next we compute the covariance matrix of W, namely $V^* = \text{cov}(W)$, where $w = (w_{11}, w_{21}, \ldots, w_{2i}, \ldots, w_{2K})$. Note that as $w_{2i} = \sum_{j \neq k}^{n} y_{ij} y_{ik}$ is the sum of products of the responses of the ith $(i = 1, \ldots, K)$ cluster and because k clusters are independent, it then follows that

$$V^* = \begin{bmatrix} \text{var}(W_{11}) & \text{cov}(W_{11}, W_{21}) & \text{cov}(W_{11}, W_{22}) & \ldots & \text{cov}(W_{11}, W_{2K}) \\ \text{cov}(W_{21}, W_{11}) & \text{var}(W_{21}) & 0 & \ldots & 0 \\ \text{cov}(W_{22}, W_{11}) & 0 & \text{var}(W_{22}) & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{cov}(W_{2K}, W_{11}) & 0 & 0 & \ldots & \text{var}(W_{2K}) \end{bmatrix},$$

$$= \begin{bmatrix} \sum_{i=1}^{K} a_i x_i^2 & b_1 x_1 & b_2 x_2 & \ldots & b_K x_K \\ b_1 x_1 & c_1 & 0 & \ldots & 0 \\ b_2 x_2 & 0 & c_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_K x_K & 0 & 0 & \ldots & c_K \end{bmatrix}, \tag{5.43}$$

where the formulas for a_i, b_i, and c_i are given by

$$a_i = \text{var}\left(\sum_{j=1}^{n} Y_{ij}\right)$$

$$= [n e_{1i} + n(n-1) e_{2i}], \tag{5.44}$$

$$b_i = \text{cov}\left(\sum_{j=1}^{n} Y_{ij}, \sum_{k \neq \ell}^{n} Y_{ik} Y_{i\ell}\right)$$

$$= 2n(n-1) e_{3i} + n(n-1)(n-2) e_{4i}, \tag{5.45}$$

and

$$c_i = \text{var}\left(\sum_{k \neq \ell}^{n} Y_{ik} Y_{i\ell}\right)$$

$$= 4\text{var}\left(\sum_{k < \ell} Y_{ik} Y_{i\ell}\right)$$

$$= 4\left[\sum_{k<\ell}^{n} \mathrm{var}(Y_{ik}Y_{i\ell}) + 2\left\{\sum_{k<\ell}\sum_{\ell<u} \mathrm{cov}(Y_{ik}Y_{i\ell}, Y_{ik}Y_{iu})\right.\right.$$

$$+\sum_{k=1}^{n-2}\sum_{\ell=k+1}^{n}\left(\sum_{u=\ell+1}^{n} \mathrm{cov}(Y_{ik}Y_{i\ell}, Y_{i\ell}Y_{iu}) + \sum_{u=k+1}^{\ell-1} \mathrm{cov}(Y_{ik}Y_{i\ell}, Y_{iu}Y_{i\ell})\right)$$

$$\left.\left. + \sum_{k<\ell}\sum_{u<v}\sum_{k\neq u}\sum_{\ell\neq v} \mathrm{cov}(Y_{ik}Y_{i\ell}, Y_{iu}Y_{iv})\right\}\right]$$

$$= 4\left[\frac{n(n-1)}{2}e_{5i} + 2\left\{\left(\sum_{j=1}^{n-2} j(j+1)/2\right)e_{6i} + \left(\sum_{j=1}^{n-2} j(j+1)\right)e_{6i}\right.\right.$$

$$\left.\left. + \{n(n-1)(n-2)(n-3)/8\}e_{7i}\right\}\right], \tag{5.46}$$

respectively, where

$$e_{1i} = E_{\gamma_i^*}h_i(\cdot) - \{E_{\gamma_i^*}h_i(\cdot)\}^2,\ e_{2i} = E_{\gamma_i^*}h_i^2(\cdot) - \{E_{\gamma_i^*}h_i(\cdot)\}^2,$$

$$e_{3i} = E_{\gamma_i^*}h_i^2(\cdot) - E_{\gamma_i^*}h_i(\cdot)E_{\gamma_i^*}h_i^2(\cdot),\ e_{4i} = E_{\gamma_i^*}h_i^3(\cdot) - E_{\gamma_i^*}h_i(\cdot)E_{\gamma_i^*}h_i^2(\cdot),$$

$$e_{5i} = E_{\gamma_i^*}h_i^2(\cdot) - \{E_{\gamma_i^*}h_i^2(\cdot)\}^2,\ e_{6i} = E_{\gamma_i^*}h_i^3(\cdot) - \{E_{\gamma_i^*}h_i^2(\cdot)\}^2, \text{ and}$$

$$e_{7i} = E_{\gamma_i^*}h_i^4(\cdot) - \{E_{\gamma_i^*}h_i^2(\cdot)\}^2,$$

with $E_{\gamma_i^*}\{h_i^r(\cdot)\} = E_{\gamma_i^*}[\exp(x_i\beta + \sigma_\gamma\gamma_i^*)/\{1+\exp(x_i\beta + \sigma_\gamma\gamma_i^*)\}]^r$, for $r = 1,2,3,4$.

Note that as $\mathrm{cov}(W_{11}, W_{2i}) \neq 0$, the V^* matrix in (5.43) is not a diagonal matrix. Consequently, it is clear from (5.42) – (5.46) that

$$B = D'V^{*-1} \neq B_0$$

given in (5.41), which contradicts the claim by Jiang and Zhang (2001) that the optimal choice of the $B = DV^{*-1}$ matrix has the simple block diagonal form, namely, $B = B_0 = \mathrm{diag}(I_1, 1_K]$, free from β and σ_γ^2. In fact, unlike Jiang and Zhang (2001), the above calculations show that the so-called improved moment estimation requires the computations for the third− and the fourth-order moments of the responses in a cluster, to construct the $\mathrm{cov}(W) = V^*$ matrix in particular.

5.2.2.2 Effect of Mis-specification For Optimal Choice

In this section, we proceed as in Jiang and Zhang (2001) and solve (5.40) for $\theta = (\beta, \sigma_\gamma)'$ with B_0 given as in (5.41), even though this B_0 is no longer an optimal substitute of the B matrix, and examine the effect of this misspecification on the asymptotic variance of the estimator of θ.

Let $\hat{\theta} = (\hat{\beta}, \hat{\sigma_\gamma})'$ be the solution of (5.40) for $\theta = (\beta, \sigma_\gamma)'$ based on Jiang and Zhang's (2001) optimal choice of $B = B_0 = \text{diag}(I_1, 1_K')$. As $E(W^*) = \psi^*$, it then follows that $\hat{\theta}$ has the asymptotic (as $M = nK \to \infty$) covariance matrix given by

$$\begin{aligned} \text{cov}(\hat{\theta}) &= (B_0 D)^{-1} B_0 E[(W^* - \psi^*)(W^* - \psi^*)'] B_0' \{(B_0 D)^{-1}\}' \\ &= (B_0 D)^{-1} B_0 V^* B_0' \{(B_0 D)^{-1}\}', \end{aligned} \tag{5.47}$$

where V^* is the true covariance matrix of the base statistic W^*, given in (5.43), and D is the derivative matrix as given by (5.42). Note that the asymptotic covariance matrix in (5.47) is the true covariance of the improved moment estimator, computed based on the misspecified B_0. If one were, however, solving the estimating equation (5.37) for θ, then the improved moment estimator of θ would have the asymptotic covariance matrix given by

$$[D' V^{*-1} D]^{-1}.$$

5.2.3 *Generalized Quasi-Likelihood (GQL) Approach*

Let $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$ be the n_i binary response vector collected from n_i members of the ith $(i = 1, \ldots, K)$ family. Next, write the mean vector of y_i and its covariance matrix as

$$\begin{aligned} E[Y_i] &= \pi_i(\beta, \sigma_\gamma^2) \\ &= (\pi_{i1}(\beta, \sigma_\gamma^2), \ldots, \pi_{ij}(\beta, \sigma_\gamma^2), \ldots, \pi_{in_i}(\beta, \sigma_\gamma^2))' : n_i \times 1 \end{aligned} \tag{5.48}$$

$$\begin{aligned} \text{Cov}[Y_i] &= \Sigma_i(\beta, \sigma_\gamma^2) \\ &= (\sigma_{ijk}) = (\lambda_{ijk}(\beta, \sigma_\gamma^2) - \pi_{ij}(\beta, \sigma_\gamma^2)\pi_{ik}(\beta, \sigma_\gamma^2)) : n_i \times n_i, \end{aligned} \tag{5.49}$$

where $\pi_{ij}(\beta, \sigma_\gamma^2)$ and $\lambda_{ijk}(\beta, \sigma_\gamma^2)$ are given in (5.7) and (5.8), respectively.

5.2.3.1 Marginal Generalized Quasi-Likelihood Estimation of β

Recall from (5.20) and (5.21) that $\pi_{ij}^{(s)}(\beta, \sigma_\gamma^2)$ and $\lambda_{ijk}^{(s)}(\beta, \sigma_\gamma^2)$ are the simulation based computational formulas for $\pi_{ij}(\beta, \sigma_\gamma^2)$ and $\lambda_{ijk}(\beta, \sigma_\gamma^2)$, respectively. We use these computational formulas and write

$$\pi_i^{(s)}(\beta, \sigma_\gamma^2) = (\pi_{i1}^{(s)}(\beta, \sigma_\gamma^2), \ldots, \pi_{ij}^{(s)}(\beta, \sigma_\gamma^2), \ldots, \pi_{in_i}^{(s)}(\beta, \sigma_\gamma^2))' : n_i \times 1 \tag{5.50}$$

$$\Sigma_i^{(s)}(\beta, \sigma_\gamma^2) = (\lambda_{ijk}^{(s)}(\beta, \sigma_\gamma^2) - \pi_{ij}^{(s)}(\beta, \sigma_\gamma^2)\pi_{ik}^{(s)}(\beta, \sigma_\gamma^2)) : n_i \times n_i, \tag{5.51}$$

as the computational formulas for $\pi_i(\beta,\sigma_\gamma^2)$ and $\Sigma_i(\beta,\sigma_\gamma^2)$ given in (5.48) and (5.49), respectively.

In the manner similar to that of the Poisson mixed model case (4.62), for given σ_γ^2, one may obtain the GQL estimate of β by solving the estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \pi_i^{(s)'}}{\partial \beta} \Sigma_i^{(s)^{-1}} (y_i - \pi_i^{(s)}) = 0, \tag{5.52}$$

[Sutradhar (2003, Section 3)] where the derivative matrix $\partial \pi_i^{(s)'}/\partial \beta$ can be computed by using the formula for $\partial \pi_{ij}^{(s)}/\partial \beta$ from (5.32), for all $j = 1,\ldots,n_i$. Let $\hat{\beta}_{GQL}$ be the solution of (5.52). This GQL estimator is consistent and highly efficient. It also follows that asymptotically (as $K \to \infty$), for known σ_γ^2, $\hat{\beta}_{GQL}$ follows the multivariate Gaussian distribution with mean β and the covariance matrix given by

$$\begin{aligned} \mathrm{cov}(\hat{\beta}_{GQL}) &= \mathrm{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial \pi_i'}{\partial \beta} \Sigma_i^{-1} \frac{\partial \pi_i}{\partial \beta'} \right]^{-1} \\ &\simeq \mathrm{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial \pi_i^{(s)'}}{\partial \beta} \Sigma_i^{(s)^{-1}} \frac{\partial \pi_i^{(s)}}{\partial \beta'} \right]^{-1}. \end{aligned} \tag{5.53}$$

5.2.3.2 Marginal Generalized Quasi-Likelihood Estimation of σ_γ

The squared binary responses and the first-order binary responses provide the same information, and because the first-order responses were used to construct the marginal GQL estimating equation for β, unlike the Poisson mixed model case (4.69) we therefore now use only second-order pairwise responses to construct the marginal GQL estimating equation for σ_γ. Thus, we write the second-order pairwise products based GQL estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma_\gamma} \Omega_i^{-1} (u_i - \lambda_i) = 0, \tag{5.54}$$

where

$$\begin{aligned} u_i &= (y_{i1}y_{i2},\ldots,y_{ij}y_{ik},\ldots,y_{i(n_i-1)}y_{in_i})' \\ \lambda_i &= E[U_i] = (\lambda_{i12},\ldots,\lambda_{ijk},\ldots,\lambda_{i(n_i-1)n_i})', \end{aligned} \tag{5.55}$$

with $\lambda_{ijk} \equiv \lambda_{ijk}(\beta,\sigma_\gamma^2)$ as given in (5.15). Furthermore, in (5.54), $\Omega_i = \mathrm{cov}[U_i]$. The formulas for the elements of this matrix can be computed as follows.

Formula for cov$[U_i] = \Omega_i$

$$\begin{aligned}
\mathrm{var}[Y_{ij}Y_{ik}] &= E[Y_{ij}^2 Y_{ik}^2] - [E[Y_{ij}Y_{ik}]]^2 \\
&= E[Y_{ij}Y_{ik}] - [E[Y_{ij}Y_{ik}]]^2 \\
&= \lambda_{ijk}(\beta,\sigma_\gamma^2)[1-\lambda_{ijk}(\beta,\sigma_\gamma^2)],
\end{aligned} \tag{5.56}$$

where $\lambda_{ijk}(\beta,\sigma_\gamma^2)$ is defined in (5.15). Similarly, we obtain

$$\mathrm{cov}[Y_{ij}Y_{ik}, Y_{i\ell}Y_{im}] = \begin{cases}
\delta_{ijkm}(\beta,\sigma_\gamma^2) - \lambda_{ijk}(\beta,\sigma_\gamma^2)\lambda_{ijm}(\beta,\sigma_\gamma^2) & \text{for } j=\ell \\
\delta_{ijk\ell}(\beta,\sigma_\gamma^2) - \lambda_{ijk}(\beta,\sigma_\gamma^2)\lambda_{ij\ell}(\beta,\sigma_\gamma^2) & \text{for } j=m \\
\delta_{ijkm}(\beta,\sigma_\gamma^2) - \lambda_{ijk}(\beta,\sigma_\gamma^2)\lambda_{ikm}(\beta,\sigma_\gamma^2) & \text{for } k=\ell \\
\delta_{ijk\ell}(\beta,\sigma_\gamma^2) - \lambda_{ijk}(\beta,\sigma_\gamma^2)\lambda_{ik\ell}(\beta,\sigma_\gamma^2) & \text{for } k=m
\end{cases}, \tag{5.57}$$

where the third-order moment $\delta_{ijkm}(\beta,\sigma_\gamma^2)$, for example, is given by (5.16). Next,

$$\mathrm{cov}[Y_{ij}Y_{ik}, Y_{i\ell}Y_{im}] = \phi_{ijk\ell m}(\beta,\sigma_\gamma^2) - \lambda_{ijk}(\beta,\sigma_\gamma^2)\lambda_{i\ell m}(\beta,\sigma_\gamma^2) \ \text{ for } j\neq\ell,\ k\neq m, \tag{5.58}$$

where the fourth-order moment $\phi_{ijk\ell m}(\beta,\sigma_\gamma^2)$ is given by (5.17).

Note that the computational formulas for

$$\lambda_{ijk}(\beta,\sigma_\gamma^2), \quad \delta_{ijkm}(\beta,\sigma_\gamma^2), \text{ and } \phi_{ijk\ell m}(\beta,\sigma_\gamma^2),$$

are given by

$$\lambda_{ijk}^{(s)}(\beta,\sigma_\gamma^2), \quad \delta_{ijkm}^{(s)}(\beta,\sigma_\gamma^2), \text{ and } \phi_{ijk\ell m}^{(s)}(\beta,\sigma_\gamma^2),$$

in (5.21), (5.22), and (5.23), respectively. We use these computational formulas in λ_i and Ω_i in (5.54) and construct the corresponding vector and matrix as $\lambda_i^{(s)}$ and $\Omega_i^{(s)}$. These substitutions lead to the computational GQL estimating equation for σ_γ as

$$\sum_{i=1}^{K} \frac{\partial \lambda_i^{(s)'}}{\partial \sigma_\gamma} \Omega_i^{(s)^{-1}} (u_i - \lambda_i^{(s)}) = 0. \tag{5.59}$$

Let $\hat{\sigma}_{\gamma,GQL}$ denote the solution of (5.59). It can be shown that asymptotically (as $K \to \infty$), for known β, the final GQL estimator obtained from (5.59) follows the univariate Gaussian distribution with mean σ_γ and the variance given by

$$\begin{aligned}
\mathrm{var}(\hat{\sigma}_{\gamma,GQL}) &= \mathrm{limit}_{K\to\infty}\left[\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \sigma_\gamma}\Omega_i^{-1}\frac{\partial \lambda_i}{\partial \sigma_\gamma}\right]^{-1} \\
&\simeq \mathrm{limit}_{K\to\infty}\left[\sum_{i=1}^{K} \frac{\partial \lambda_i^{(s)'}}{\partial \sigma_\gamma}\Omega_i^{(s)^{-1}}\frac{\partial \lambda_i^{(s)}}{\partial \sigma_\gamma}\right]^{-1},
\end{aligned} \tag{5.60}$$

where the formula for the computation of $\partial \lambda_i^{(s)'}/\partial \sigma_\gamma$ is available from (5.35).

Note that in practice, the iterative equations (5.52) for β and (5.59) for σ_γ constitute a cycle, and the cycles of operation continue until convergence, to obtain the final GQL estimates $\hat{\beta}_{GQL}$ and $\hat{\sigma}_{\gamma,GQL}$ for β and σ_γ, respectively.

5.2.3.3 Joint Generalized Quasi-Likelihood (GQL) Estimation for β and σ_γ

For quick convergence of the estimates, one may like to estimate β and σ_γ jointly. For this, the estimating equations (5.52) and (5.54) may be combined as follows. Let

$$s_i = (y_i', u_i')',$$

where

$$y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})', \text{ and } u_i = (y_{i1}y_{i2}, \ldots, y_{ij}y_{ik}, \ldots, y_{i(n_i-1)}y_{in_i})'.$$

Note that $E[Y_i] = \pi_i(\beta, \sigma_\gamma^2)$ as given in (5.48), and $E[U_i] = \lambda_i(\beta, \sigma_\gamma^2)$ as given by (5.55). Thus, we write

$$E[S_i] = \zeta_i(\beta, \sigma_\gamma^2) = (\pi_i', \lambda_i')'. \tag{5.61}$$

Furthermore, let

$$\text{cov}[S_i] = \Upsilon_i = \begin{bmatrix} \text{cov}(Y_i) & \text{cov}(Y_i, U_i') \\ & \text{cov}(U_i) \end{bmatrix} \tag{5.62}$$

$$= \begin{bmatrix} \Sigma_i & \Lambda_i \\ & \Omega_i \end{bmatrix}, \tag{5.63}$$

where $\Sigma_i = \text{cov}(Y_i)$ is given by (5.49), and $\Omega_i = \text{cov}(U_i)$ is constructed by using the formulas from (5.56) to (5.58). To construct Υ_i, it remains to compute the elements of the $\Lambda_i = \text{cov}(Y_i, U_i')$ matrix. In the present binary mixed model case, this can be done as follows.

Formula for cov$[Y_i, U_i'] = \Lambda_i$

$$\text{cov}[Y_{ij}, Y_{ik}Y_{i\ell}] = \begin{cases} E[Y_{ij}^2 Y_{i\ell}] - \pi_{ij}\lambda_{ij\ell} = \lambda_{ij\ell} - \pi_{ij}\lambda_{ij\ell} & \text{for } j = k \\ E[Y_{ij}^2 Y_{ik}] - \pi_{ij}\lambda_{ijk} = \lambda_{ijk} - \pi_{ij}\lambda_{ijk} & \text{for } j = \ell \\ E[Y_{ij}Y_{ik}Y_{i\ell}] - \pi_{ij}\lambda_{ik\ell} = \delta_{ijk\ell} - \pi_{ij}\lambda_{ik\ell} & \text{for } j \neq k; j \neq \ell, \end{cases} \tag{5.64}$$

where $\pi_{ij}(\beta, \sigma_\gamma^2)$ and $\lambda_{ijk}(\beta, \sigma_\gamma^2)$, are given by (5.7) and (5.8), respectively. The formula for the third-order moment $\delta_{ijk\ell}(\beta, \sigma_\gamma^2)$ is given in (5.16).

Next, for $\theta = (\beta', \sigma_\gamma)'$, in the manner similar to that of (5.52) and (5.59), one may construct the GQL estimating equation for θ given by

$$\sum_{i=1}^{K} \frac{\partial \zeta_i^{(s)'}}{\partial \theta} \Upsilon_i^{(s)^{-1}} (s_i - \zeta_i^{(s)}) = 0, \tag{5.65}$$

where $\zeta_i^{(s)}$ is computed from ζ_i by replacing $\pi_{ij}(\beta, \sigma_\gamma^2)$ and $\lambda_{ijk}(\beta, \sigma_\gamma^2)$, with $\pi_{ij}^{(s)}(\beta, \sigma_\gamma^2)$ and $\lambda_{ijk}^{(s)}(\beta, \sigma_\gamma^2)$, respectively. Similarly, $\Upsilon_i^{(s)}$ is constructed from Υ_i by replacing

$$\pi_{ij}(\beta, \sigma_\gamma^2),\ \lambda_{ijk}(\beta, \sigma_\gamma^2),\ \text{and}\ \delta_{ijk\ell}(\beta, \sigma_\gamma^2),$$

with

$$\pi_{ij}^{(s)}(\beta, \sigma_\gamma^2),\ \lambda_{ijk}^{(s)}(\beta, \sigma_\gamma^2),\ \text{and}\ \delta_{ijk\ell}^{(s)}(\beta, \sigma_\gamma^2),$$

respectively, where the formulas for the latter simulation based functions are given in (5.20), (5.21), and (5.22), respectively.

Note that the estimating equation (5.65) can be solved by using the iterative equation

$$\hat{\theta}_{GQL}(r+1) = \hat{\theta}_{GQL}(r) + \left[\sum_{i=1}^{K} \frac{\partial \zeta_i^{(s)'}}{\partial \theta} \Upsilon_i^{(s)^{-1}} \frac{\partial \zeta_i^{(s)}}{\partial \theta'}\right]_r^{-1} \times \left[\sum_{i=1}^{K} \frac{\partial \zeta_i^{(s)'}}{\partial \theta} \Upsilon_i^{(s)^{-1}} (s_i - \zeta_i^{(s)})\right]_r, \tag{5.66}$$

where $[]_r$ denotes that the expression within the square bracket is evaluated at $\theta = \hat{\theta}_{GQL}(r)$, the estimate obtained for the rth iteration. Furthermore, similar to that of (5.53) or (5.60), it can be shown that asymptotically (as $K \to \infty$), the final GQL estimator obtained from (5.66) follows the multivariate Gaussian distribution with mean θ and the variance given by

$$\begin{aligned} \text{var}(\hat{\theta}_{GQL}) &= \text{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \theta} \Upsilon_i^{-1} \frac{\partial \zeta_i}{\partial \theta'}\right]^{-1} \\ &\simeq \text{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial \zeta_i^{(s)'}}{\partial \theta} \Upsilon_i^{(s)^{-1}} \frac{\partial \zeta_i^{(s)}}{\partial \theta'}\right]^{-1}. \end{aligned} \tag{5.67}$$

5.2.4 Maximum Likelihood (ML) Estimation

It follows from the model (5.1) that the likelihood function for the data is given by

$$L(\beta, \sigma_\gamma^2) = \Pi_{i=1}^{K} L_i(\beta, \sigma_\gamma^2) = \Pi_{i=1}^{K} f(y_i|\gamma_i). \tag{5.68}$$

Let $y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$. Then the log of likelihood function (5.68) may be expressed as

$$\log L = \sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}x'_{ij}\beta + \sum_{i=1}^{K} \log J_i, \tag{5.69}$$

where

$$J_i = \int_{-\infty}^{\infty} \exp(\sigma_\gamma y_{i\cdot}\gamma_i^*)\Delta_i(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^* \equiv \int_{-\infty}^{\infty} \tilde{J}_i(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^*, \tag{5.70}$$

with $\Delta_i(\gamma_i^*) = \{\Pi_{j=1}^{n_i}[1+\exp(x'_{ij}\beta + \sigma_\gamma\gamma_i^*)]\}^{-1}$. For likelihood estimation of β and σ_γ, we now consider the joint score equations $U_1(\beta,\sigma_\gamma^2) = \partial \log L/\partial\beta = 0$ and $U_2(\beta,\sigma_\gamma^2) = \partial \log L/\partial\sigma_\gamma = 0$, where

$$U_1(\beta,\sigma_\gamma^2) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\{y_{ij} - A_{ij}/J_i\}x_{ij} \equiv \sum_{i=1}^{K} U_{1i}(\beta,\sigma_\gamma^2), \tag{5.71}$$

$$U_2(\beta,\sigma_\gamma^2) = \sum_{i=1}^{K} M_i/J_i \equiv \sum_{i=1}^{K} U_{2i}(\beta,\sigma_\gamma^2), \tag{5.72}$$

with J_i as in (5.70), and

$$A_{ij} = \int_{-\infty}^{\infty} \exp\{\sigma_\gamma y_{i\cdot}\gamma_i^*\}\Delta_i(\gamma_i^*)\pi_{ij}^*(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^* \equiv \int_{-\infty}^{\infty} \tilde{A}_{ij}(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^*,$$

$$M_i = \int_{-\infty}^{\infty} \exp\{\sigma_\gamma y_{i\cdot}\gamma_i^*\}\left[\sum_{j=1}^{n_i}(y_{ij} - \pi_{ij}^*(\gamma_i^*))\right]\Delta_i(\gamma_i^*)\gamma_i^* g_N(\gamma_i^*|1)d\gamma_i^*$$
$$\equiv \int_{-\infty}^{\infty} \tilde{M}_i(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^*.$$

Next, we approximate J_i, A_{ij}, and M_i, with

$$J_i^{(s)} = \frac{1}{N}\sum_{w=1}^{N}\tilde{J}_i(\gamma_{iw}^*),\;\; A_{ij}^{(s)} = \frac{1}{N}\sum_{w=1}^{N}\tilde{A}_{ij}(\gamma_{iw}^*),\;\text{and}\;\; M_i^{(s)} = \frac{1}{N}\sum_{w=1}^{N}\tilde{M}_i(\gamma_{iw}^*), \tag{5.73}$$

respectively, and compute $U_1^{(s)}(\beta,\sigma_\gamma^2)$ and $U_2^{(s)}(\beta,\sigma_\gamma^2)$ from $U_1(\beta,\sigma_\gamma^2)$ in (5.71) and $U_2(\beta,\sigma_\gamma^2)$ in (5.72), by replacing A_{ij}, J_i, and M_i with $A_{ij}^{(s)}$, $J_i^{(s)}$, and $M_i^{(s)}$, respectively. One may now obtain the maximum likelihood estimates of β and σ_γ, by solving the estimating equations

$$U_1^{(s)}(\beta,\sigma_\gamma^2) = 0, \;\text{and}\;\; U_2^{(s)}(\beta,\sigma_\gamma^2) = 0, \tag{5.74}$$

jointly. Let $\hat{\beta}_{ML}$ and $\hat{\sigma}_{\gamma,ML}$ denote these maximum likelihood estimates for β and σ_γ, respectively.

Next, we briefly address the issue of estimating the asymptotic covariance matrix of $\hat{\theta}_{ML} = (\hat{\beta}'_{ML}, \hat{\sigma}_{\gamma,ML})'$. The asymptotic covariance matrix of $\hat{\theta}_{ML}$ is given by

$$\mathrm{cov}(\hat{\theta}_{ML}) = \mathrm{limit}_{K\to\infty}\{I(\theta)\}^{-1}, \tag{5.75}$$

where by standard regularity conditions,

$$I(\theta) = -\begin{bmatrix} V_{11} & V_{12} \\ & V_{22} \end{bmatrix} \tag{5.76}$$

with

$$V_{11} = E_y\{\partial U_1(\beta,\sigma_\gamma^2)/\partial\beta'\},\quad V_{12} = E_y\{\partial U_1(\beta,\sigma_\gamma^2)/\partial\sigma_\gamma\},$$

and

$$V_{22} = E_y\{\partial U_2(\beta,\sigma_\gamma^2)/\partial\sigma_\gamma\},$$

where $U_1(\beta,\sigma_\gamma^2)$ and $U_2(\beta,\sigma_\gamma^2)$ are the score functions as in (5.71) and (5.72), respectively.

Note that although the $I(\theta)$ matrix given by (5.76) requires the computation of the expectations of the second order derivatives, the existing computer packages such as SAS program NLMIXED (SAS/STAT User guide, Version 8, Volume 2, p. 2475 – 76), however, computes the $I(\theta)$ on the basis of the observed information matrix, which may not be reliable. Here, we compute the $I(\theta)$ matrix by using $V_{11}^{(s)}$, $V_{12}^{(s)}$, and $V_{22}^{(s)}$ for V_{11}, V_{12}, and V_{22}, respectively, where

$$V_{11}^{(s)} = \sum_{i=1}^{K}\left[\sum_{y_{i1}=0}^{1}\cdots\sum_{y_{in_i}=0}^{1}\partial\{U_{1i}^{(s)}(\beta,\sigma_\gamma^2)/\partial\beta'\}\right] \tag{5.77}$$

$$V_{12}^{(s)} = \sum_{i=1}^{K}\left[\sum_{y_{i1}=0}^{1}\cdots\sum_{y_{in_i}=0}^{1}\partial\{U_{1i}^{(s)}(\beta,\sigma_\gamma^2)/\partial\sigma_\gamma\}\right] \tag{5.78}$$

$$V_{22}^{(s)} = \sum_{i=1}^{K}\left[\sum_{y_{i1}=0}^{1}\cdots\sum_{y_{in_i}=0}^{1}\partial\{U_{2i}^{(s)}(\beta,\sigma_\gamma^2)/\partial\sigma_\gamma\}\right] \tag{5.79}$$

$$V_{21}^{(s)} = V_{12}^{(s)}, \tag{5.80}$$

where $U_{1i}^{(s)}(\beta,\sigma_\gamma^2)$ and $U_{2i}^{(s)}(\beta,\sigma_\gamma^2)$ are obtained from $U_{1i}(\beta,\sigma_\gamma^2)$ and $U_{2i}(\beta,\sigma_\gamma^2)$ in (5.71) – (5.72), by using $J_i^{(s)}$, $A_{ij}^{(s)}$, and $M_i^{(s)}$, for J_i, A_{ij}, and M_i, respectively.

5.2.5 Asymptotic Efficiency Comparison

In this section, we provide a numerical comparison among the asymptotic variances of the IMM, GQL, and ML estimators obtained from (5.40), (5.65), and (5.74), respectively. The formulas for the asymptotic covariance matrices for these three estimators are, respectively, given by (5.47), (5.67), and (5.75). As far as the model is concerned, we have chosen, for example, the same simple binary logistic model considered in Section 5.2.2.1 [see also Jiang and Zhang (2001, 756 – 57)]. For convenience, we rewrite this model here as

$$\text{logit}\{\Pr(y_{ij}=1|\gamma_i^*)\} = x_i\beta + \sigma_\gamma \gamma_i^*, \tag{5.81}$$

where $\gamma_i^* \stackrel{iid}{\sim} N(0,1)$. As in practice cluster sizes are usually small, for example, in a familial mixed model, family size can be 3 or 4, we choose $n_i = n = 4$, that is, families each with 4 members, for example. For parameter values, we consider, for example, two values for the scalar β, namely,

$$\beta = 0.0 \text{ and } 1.0$$

and some small and large values for σ_γ as

$$\sigma_\gamma = 0.5, 0.75, 1.0, 1.25, \text{ and } 1.5.$$

As far as the design covariate x_i is concerned, we choose

$$x_i = \begin{cases} -1.0, & \text{for } i = 1,\ldots,K/4 \\ 0.0, & \text{for } i = K/4+1,\ldots,3K/4 \\ 1.0 & \text{for } i = 3K/4+1,\ldots,K. \end{cases} \tag{5.82}$$

For family number, we choose $K = 500$.

5.2.5.1 Asymptotic variance of the IMM Estimator

To compute the asymptotic covariance of the IMM estimator of $\theta = (\beta, \sigma_\gamma)'$ by (5.47), that is,

$$\text{cov}[\hat{\theta}_{IMM}] = (B_0 D)^{-1} B_0 V^* B_0' \{(B_0 D)^{-1}\}' \tag{5.83}$$

under the model (5.81), with

$$B_0 = \begin{bmatrix} 1\ 0 \ldots 0 \\ \\ 0\ 1 \ldots 1 \end{bmatrix},$$

as in Jiang and Zhang (2001), we can use the formulas for D and V^* matrices from (5.42) and (5.43), and obtain

$$B_0 D = n \begin{bmatrix} \sum_{i=1}^{K} x_i e_{1i}^* & \sum_{i=1}^{K} x_i e_{2i}^* \\ 2(n-1)\sum_{i=1}^{K} e_{3i}^* & 2(n-1)\sum_{i=1}^{K} e_{4i}^* \end{bmatrix}, \tag{5.84}$$

and

$$B_0 V^* B_0' = \begin{bmatrix} \sum_{i=1}^{K} a_i x_i^2 & \sum_{i=1}^{K} b_i x_i \\ \sum_{i=1}^{K} b_i x_i & \sum_{i=1}^{K} c_i \end{bmatrix}. \tag{5.85}$$

Note that the formulas for e_{ri}^* for $r = 1, \ldots, 4$; and for a_i, b_i, and c_i, in terms of e_{ri} for $r = 1, \ldots, 7$; are available from (5.42) and (5.44) – (5.46). For $n = 4$, it follows that the formulas for a_i, b_i, and c_i reduce to

$$a_i = 4e_{1i} + 12e_{2i},\ b_i = 24[e_{3i} + e_{4i}],\ \text{and}\ c_i = 4[6e_{5i} + 24e_{6i} + 6e_{7i}].$$

Further note that the computations for $e_{1i}, \ldots, e_{7i}$, and $e_{1i}^*, \ldots, e_{4i}^*$ necessary for the aforementioned computations are done by simulating N values of γ_i^* only once from the $N(0,1)$ distribution. Here we choose $N = 5000$, a sufficiently large value for the approximations of the expectations involved in the formulas for $e_{1i}, \ldots, e_{7i}, e_{1i}^*, \ldots, e_{4i}^*$. Thus, e_{1i}^*, for example, is approximated by

$$e^{*(s)}_{1i} = (N)^{-1} \sum_{w=1}^{N} [h_i(\gamma_{iw}^*)\{1.0 - h_i(\gamma_{iw}^*)\}]x_i$$

and similarly, e_{3i}, for example, is approximated by

$$e_{3i}^{(s)} = [(N)^{-1} \sum_{w=1}^{N} h_i^2(\gamma_{iw}^*)][1 - (N)^{-1} \sum_{w=1}^{N} h_i(\gamma_{iw}^*)],$$

where

$$h_i(\gamma_{iw}^*) = \frac{exp(x_i\beta + \sigma_\gamma \gamma_{iw}^*)}{1 + exp(x_i\beta + \sigma_\gamma \gamma_{iw}^*)}.$$

The covariance of $\hat{\theta}_{IMM}$ is now immediate by using (5.84) and (5.85) in (5.83).

5.2.5.2 Asymptotic Variance of the GQL Estimator

The formula for the asymptotic covariance matrix of $\hat{\theta}_{GQL}$ is given by (5.67), that is,

$$\text{cov}(\hat{\theta}_{GQL}) = \text{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial \zeta_i^{(s)'}}{\partial \theta} \Upsilon_i^{(s)^{-1}} \frac{\partial \zeta_i^{(s)}}{\partial \theta'} \right]^{-1}. \tag{5.86}$$

To compute this covariance matrix, we need to compute the submatrices in (5.86) under the model (5.81), for $n = 4$.

First, the formula for the derivative matrix $\partial \zeta_i' / \partial \theta$ is given by

$$\frac{\partial \zeta_i'}{\partial \theta} = \begin{bmatrix} 1_4' e_{1i}^* \; 2(1_6' e_{3i}^*) \\ \\ 1_4' e_{2i}^* \; 2(1_6' e_{4i}^*) \end{bmatrix}, \tag{5.87}$$

where e_{1i}^*, e_{2i}^*, e_{3i}^*, and e_{4i}^* are defined as in (5.42). Next, recall from (5.63) that

$$\Upsilon_i = \begin{bmatrix} \Sigma_i \; \Lambda_i \\ \\ \Lambda_i' \; \Omega_i \end{bmatrix}. \tag{5.88}$$

For the present special case, that is, under the model (5.81), with $n = 4$, the submatrices in (5.88) can be computed as

$$\Sigma_i = \begin{bmatrix} e_{i1} & e_{2i} & e_{2i} & e_{2i} \\ & e_{1i} & e_{2i} & e_{2i} \\ & & e_{1i} & e_{2i} \\ & & & e_1 \end{bmatrix} \tag{5.89}$$

$$\Lambda_i = \begin{bmatrix} e_{3i} & e_{3i} & e_{3i} & e_{4i} & e_{4i} & e_{4i} \\ e_{3i} & e_{4i} & e_{4i} & e_{3i} & e_{3i} & e_{4i} \\ e_{4i} & e_{3i} & e_{4i} & e_{3i} & e_{4i} & e_{3i} \\ e_{4i} & e_{4i} & e_{3i} & e_{4i} & e_{3i} & e_{3i} \end{bmatrix}, \tag{5.90}$$

$$\Omega_i = \begin{bmatrix} e_{5i} & e_{6i} & e_{6i} & e_{6i} & e_{6i} & e_{7i} \\ & e_{5i} & e_{6i} & e_{6i} & e_{7i} & e_{6i} \\ & & e_{5i} & e_{7i} & e_{6i} & e_{6i} \\ & & & e_{5i} & e_{6i} & e_{6i} \\ & & & & e_{5i} & e_{6i} \\ & & & & & e_{5i} \end{bmatrix}, \tag{5.91}$$

respectively, where $e_{\ell i}$ for $\ell = 1,\ldots,7$ are given as in (5.44) – (5.46).

Now by approximating e_{ri}^* $(r = 1,\ldots,4)$ and $e_{\ell i}$ $(\ell = 1,\ldots,7)$, by $e^{*(s)}_{ri}$ and $e_{\ell i}^{(s)}$, respectively, one may compute the $\text{cov}(\hat{\theta}_{GQL})$ for $n = 4$ by using the formulas from (5.87) and (5.88) in (5.86).

5.2.5.3 Asymptotic Variance of the ML Estimator

By (5.75) and (5.76), we write the asymptotic covariance matrix of $\hat{\theta}_{ML}$, as

$$\text{cov}[\hat{\theta}_{ML}] = -\begin{bmatrix} V_{11}^{(s)} & V_{12}^{(s)} \\ & V_{22}^{(s)} \end{bmatrix}^{-1}, \tag{5.92}$$

where $V_{11}^{(s)}$, $V_{12}^{(s)}$, and $V_{22}^{(s)}$, are given by (5.77), (5.78), and (5.79), respectively. To compute this second-order derivative matrix, for convenience, we rewrite the first-order derivatives from (5.71) and (5.72), as

$$U_1(\beta, \sigma_\gamma^2) = \sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}x_{ij} + \sum_{i=1}^{K}\frac{J_{i\beta}}{J_i}, \text{ and } U_2(\beta, \sigma_\gamma^2) = \sum_{i=1}^{K}\frac{J_{i\sigma_\gamma}}{J_i},$$

respectively, where

$$J_{i\beta} = \frac{\partial J_i}{\partial \beta}, \text{ and } J_{i\sigma_\gamma} = \frac{\partial J_i}{\partial \sigma_\gamma}.$$

It then follows that

$$\begin{aligned} V_{11}^{(s)} &= \sum_{i=1}^{K} E_{y_{i1},\ldots,y_{in_i}}\left[\{J_i^{(s)}J_{i\beta\beta}^{(s)} - \{J_{i\beta}^{(s)}\}^2\}/(J_i^{(s)})^2\right] \\ &= \sum_{i=1}^{K}\sum_{yi1=0}^{1}\cdots\sum_{y_{in_i}=0}^{1}\left[\{J_i^{(s)}J_{i\beta\beta}^{(s)} - \{J_{i\beta}^{(s)}\}^2\}/(J_i^{(s)})^2\right] \end{aligned} \tag{5.93}$$

$$\begin{aligned} V_{12}^{(s)} &= \sum_{i=1}^{K} E_{y_{i1},\ldots,y_{in_i}}\left[\{J_i^{(s)}J_{i\beta\sigma_\gamma}^{(s)} - J_{i\beta}^{(s)}J_{i,\sigma_\gamma}^{(s)}]/(J_i^{(s)})^2\right] \\ &= \sum_{i=1}^{K}\sum_{yi1=0}^{1}\cdots\sum_{y_{in_i}=0}^{1}\left[\{J_i^{(s)}J_{i\beta\sigma_\gamma}^{(s)} - J_{i\beta}^{(s)}J_{i,\sigma_\gamma}^{(s)}]/(J_i^{(s)})^2\right] \end{aligned} \tag{5.94}$$

$$\begin{aligned} V_{22}^{(s)} &= \sum_{i=1}^{K} E_{y_{i1},\ldots,y_{in_i}}\left[\{J_i^{(s)}J_{i\sigma_\gamma\sigma_\gamma}^{(s)} - \{J_{i\sigma_\gamma}^{(s)}\}^2\}/(J_i^{(s)})^2\right], \\ &= \sum_{i=1}^{K}\sum_{yi1=0}^{1}\cdots\sum_{y_{in_i}=0}^{1}\left[\{J_i^{(s)}J_{i\sigma_\gamma\sigma_\gamma}^{(s)} - \{J_{i\sigma_\gamma}^{(s)}\}^2\}/(J_i^{(s)})^2\right], \end{aligned} \tag{5.95}$$

where

$$J_{i\beta\beta} = \frac{\partial J_{i\beta}}{\partial \beta'}, \quad J_{i\beta\sigma_\gamma} = \frac{\partial J_{i\beta}}{\partial \sigma_\gamma}, \text{ and } J_{i\sigma_\gamma\sigma_\gamma} = \frac{\partial J_{i\sigma_\gamma}}{\partial \sigma_\gamma}.$$

Note that for the current special case $n_i = 4$, and β is a scalar parameter. The second-order derivative elements in (5.93) – (5.95) may easily be calculated (see Exercise 5.2) under the simple logistic model (5.81). One now uses these elements in (5.92) to obtain the covariance matrix of $\hat{\theta}_{ML}$.

5.2.5.4 Numerical Comparison

For $K = 200$ families each containing $n_i = n = 4$ members with covariates given as in (5.82), we now compute the asymptotic variances for the estimators of β and σ_γ by using (5.83), (5.86), and (5.92), under the IMM, GQL, and ML approaches, respectively. For selected parameter values, these asymptotic variances are shown in Table 5.1.

Table 5.1 Comparison of asymptotic variances of the IMM, GQL, and ML estimators for the estimation of regression (β) and variance component (σ_γ) parameters of a simple binary logistic mixed model, with $n_i = n = 4$ $(i = 1, \ldots, K)$ for K=200.

Regression			Asymptotic Variances					
Parameter (β)	Method	Quantity	σ_γ= 0.25	0.50	0.75	1.00	1.25	1.50
0.0	ML	var$(\hat{\beta})$	0.0007	0.0008	0.0010	0.0014	0.0017	0.0022
		var$(\hat{\sigma}_\gamma)$	0.0003	0.0003	0.0005	0.0007	0.0013	0.0028
	GQL	var$(\hat{\beta})$	0.0108	0.0130	0.0168	0.0220	0.0288	0.0372
		var$(\hat{\sigma}_\gamma)$	0.0664	0.0259	0.0205	0.0213	0.0248	0.0305
	IMM	var$(\hat{\beta})$	0.0108	0.0130	0.0168	0.0220	0.0288	0.0372
		var$(\hat{\sigma}_\gamma)$	0.4340	0.1684	0.1320	0.1379	0.1646	0.2090
1.0	ML	var$(\hat{\beta})$	0.0046	0.0028	0.0032	0.0038	0.0045	0.0055
		var$(\hat{\sigma}_\gamma)$	0.0134	0.0019	0.0014	0.0016	0.0022	0.0030
	GQL	var$(\hat{\beta})$	0.0143	0.0167	0.0205	0.0260	0.0331	0.0418
		var$(\hat{\sigma}_\gamma)$	0.0802	0.0299	0.0228	0.0230	0.0264	0.0320
	IMM	var$(\hat{\beta})$	0.0201	0.0237	0.0293	0.0373	0.0477	0.0604
		var$(\hat{\sigma}_\gamma)$	0.5664	0.2078	0.1539	0.1537	0.1778	0.2211

It is clear from the results of Table 5.1 that the ML approach produces the estimates for both β and σ_γ parameters with uniformly smaller variances than the GQL and IMM approaches, as expected. For example, when $\beta = 1.0$ and $\sigma_\gamma = 1.25$, the ML, GQL, and IMM approaches estimate β with variances:

$$0.0045,\ 0.0331,\ \text{and}\ 0477,$$

respectively, and they estimate σ_γ with variances:

$$0.0022,\ 0.0264,\ \text{and}\ 0.1778,$$

respectively. When the GQL and IMM approaches are compared, both of these approaches appear to estimate β with almost equal variances, the GQL being slightly better, but for the estimation of σ_γ, the GQL approach is much more efficient than the IMM approach. The IMM approach always produces σ_γ estimates with very large variance. Thus, this IMM approach cannot be trusted for the estimation of the variance component of the random effects. Between the ML and GQL approaches, even though the ML approach is always better than the GQL approach, the variances of the estimators under these two approaches are not too different. Thus, the

GQL approach is highly competitive to the optimal ML approach for the estimation of both parameters. Note, however, that the ML approach is computationally cumbersome as compared to the GQL approach. Further note that in the longitudinal setup, discussed in the next two chapters, it is either impossible or extremely complicated to obtain the ML estimates. This makes the GQL an unified highly efficient estimation approach in both familial and longitudinal setups.

5.2.6 COPD Data Analysis: A Numerical Illustration

Consider the chronic obstructive pulmonary disease data, previously analyzed by Cohen (1980), Liang, Zeger, and Qaqish (1992), and Ekholm, Smith, and McDonald (1995), among others. This dataset contains the IPF (impaired pulmonary function) status of 203 siblings of 100 COPD patients, along with the information of their covariates sex, race, age, and smoking status. The IPF status was coded as 0 for a sibling with IPF, and 1 for a sibling without IPF. The complete COPD data along with covariate information is found in the appendix in Tables 5A to 5E [see also Liang et al (1992)], where Table 5A contains the COPD data from the siblings of 48 patients each with one sibling. Table 5B similarly contains the data from the siblings of COPD patients each with two siblings, and so on. The distribution of the COPD patients with their sibling sizes is in Table 5.2.

Table 5.2 Summary statistics of COPD patients and their siblings.

	Sibling Size				
	1	2	3	4	6
No. of COPD patients	48	23	17	7	5
Total siblings	48	46	51	28	30

It is of scientific interest to investigate the effects of the covariates: sex, race, age, and smoking status on the IPF status of the siblings of a COPD patient, after taking the familial correlations among the responses from the siblings of the same patient, into account.

Note that Liang, Zeger, and Qaqish (1992) used the fixed binary logistic model to analyze these COPD data. More specifically, they assume that the probability that a sibling had IPF satisfies

$$\text{logit}[Pr(y_{ij} = 1)] = x_{ij}^T \beta, \tag{5.96}$$

where for $p = 5$, $\beta = (\beta_0, \beta_1, \ldots, \beta_4)^T$ with β_0 as the intercept and remaining four components of the β vector representing the effects of the four covariates: sex, race, smoking status, and age, respectively. As the binary responses of the siblings of a COPD patient or family are likely to be correlated because of the common family effect shared by the siblings, Liang, Zeger, and Qaqish (1992) [see also Ekholm et al. (1995)] modelled these correlations through pairwise odds ratios across the families with more than one sibling. But as it is clear from Table 5A (see also Table 5.2) that 48 COPD patients have one sibling each, the pairwise odds ratio approach does not appear to address the issue of family effects properly. This is because, when there are at least two siblings in a family, the responses of these siblings get correlated as they share the common family effect. It does not mean that there is no family effect on the response of the only sibling in the family. Furthermore, the modelling of the pairwise odds ratios [of the form $\exp(\alpha)$] considered by Liang, Zeger, and Qaqish (1992) seems to be arbitrary.

Ekholm, Smith and McDonald (1995) have also analyzed these COPD data. These authors, unlike Liang, Zeger, and Qaqish (1992) developed a multivariate binary distribution by modelling the association using certain dependence ratios defined in terms of the mean parameters. This permits flexible modelling of higher-order associations, using maximum likelihood estimation. Note, however, that as there may be many higher-order association parameters depending on the cluster size, Ekholm et al. (1995) assumed a homogeneous association structure in analyzing the COPD data, mainly for the reduction of the number of association parameters, as the estimation of parameters becomes complicated without such assumptions. This assumption of homogeneous association structure also appears to be arbitrary.

Sutradhar and Mukerjee (2005), unlike Liang, Zeger, and Qaqish (1992) and Ekholm, Smith, and McDonald (1995), have fitted the binary mixed model $(5.1)-(5.3)$ to the COPD data. This allows the responses of the siblings in a family of size more than one to be overdispersed as well as correlated through the random effect of the family which is shared by all siblings. For the COPD patient with one sibling, the patient/family effect would cause the overdispersion in the binary responses. Furthermore, the binary mixed model $(5.1)-(5.3)$ would naturally accommodate the higher-order moments or correlations, as unconditionally, the responses will have an implicit joint probability distribution.

Note that σ_γ^2 in (5.3) denotes the variance among the unobserved family effects, that is, among 100 patients for the COPD data. This parameter representing the patient effect influences the unconditional mean and variance $(5.7)-(5.8)$ of each sibling of a patient and it also affects the correlations (5.10) of the responses from the siblings of the same patient. In $(5.7)-(5.8)$, or equivalently in the model $(5.1)-(5.3)$, β for the COPD data represents the effects of the covariates of the siblings on their IPF status. In notation, these covariates are: gender $(x_{ij1})[GR]$, race (x_{ij2})[RC], smoking status (x_{ij3})[SMO], and age of the sibling (x_{ij4}); and they are coded as follows:

$$x_{ij1} = \begin{cases} 0 & \text{female} \\ 1 & \text{male} \end{cases} \qquad x_{ij2} = \begin{cases} 0 & \text{white} \\ 1 & \text{black} \end{cases}$$

$$x_{ij3} = \begin{cases} 0 & \text{nonsmoker} \\ 1 & \text{smokes} \end{cases} \qquad x_{ij4} = \text{ exact age of the individual.}$$

We have also considered an intercept parameter β_0 [INTC], and hence the linear predictor in (5.3) has the form:

$$\eta_{ij}(\beta, \sigma_\gamma, \gamma_i^*) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \beta_4 x_{ij4} + \sigma_\gamma \gamma_i^*,$$

leading the probability that a sibling has the IPF disease, as

$$Pr[Y_{ij} = 1|\gamma_i^*] = \frac{\exp[\eta_{ij}(\beta, \sigma_\gamma, \gamma_i^*)]}{1 + \eta_{ij}(\beta, \sigma_\gamma, \gamma_i^*)}.$$

The regression parameters vector $\beta = [\beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4]'$ and the random effects (of the patients) variance component σ_γ were estimated by using the ML approach discussed in Section 5.2.4. Note that to compute the ML estimates we have used the MM estimates as initial values for the parameters while solving the ML estimating equations iteratively. For convenience, these MM estimates along with the final ML estimates are given in Table 5.3. Note that there was no reason to include the IMM approach in the present analysis as it was shown in the last section that this approach is asymptotically less efficient than the GQL approach. This, however, indicates that we could consider the GQL approach in the analysis, but it was also not exploited. This is because our purpose here is to demonstrate that even though in general the ML approach is complicated, in the present binary mixed model setup this approach is, however, manageable.

Table 5.3 Estimates of regression and random family (patient) effects' variance parameter and their estimated standard errors for the COPD data.

		Parameters					
Method		INTC	GR	RC	SMO	Age	σ_γ
ML	Estimate	-0.770	-0.802	-0.729	1.007	0.041	1.030
	ESE	0.136	0.141	0.191	0.162	0.001	0.494
MM	Estimate	-1.653	-0.994	-1.072	1.388	0.075	0.902
	ESE	0.182	0.230	0.251	0.216	0.001	0.627

We now interpret the ML estimates. First, the high value (1.007) of $\hat{\beta}_{ML,3}$ (smoking effect) reveals that smoking has a detrimental effect on the IPF of the siblings. Furthermore, as the gender and race were coded as 1 for male and black, respectively, the negative values of $\hat{\beta}_{ML,1}$ and $\hat{\beta}_{ML,2}$, respectively, indicate that males and black are at more risk of IPF as compared to females and individuals from white race. Similarly, the positive value of $\hat{\beta}_{ML,4}$ indicates that as age increases the risk of IPF increases too, as expected.

As far as the variance component of the random effects of the COPD patients is concerned, the estimate for σ_γ appears to be quite large, indicating that the familial correlations cannot be ignored in any inferences for the COPD data. That is, if the familial correlations are ignored, then one would obtain a misleading mean (unconditional probability for IPF status) and variance of the data.

5.3 Binary Mixed Models with Multidimensional Random Effects

Recall from Chapter 4, more specifically from Section 4.3 that there are situations where familial data may be influenced by multiple random effects. For example, (1) it was indicated in Section 4.3.1 that the responses of the jth member of the ith family may be influenced by random family effects as well as by random member (within a family) effects. Next it was indicated in Section 4.3.2 that (2) there can be a random family effect for a group of families implying that the whole dataset may contain multiple random effects depending on the number of groups. Furthermore, Section 4.3.3 provides details on the inferences for mixed models for count data, where (3) the response of a member in a given family is influenced by multiple random family effects that arise from independent sources.

In this section we deal with a binary mixed model with multiple random effects arising in a multiway factorial design setup. This factorial design set up is similar but different from the familial setup with multiple random effects considered in Section 4.3.1. For simplicity we consider two random effects arising in a two-way factorial design setup. We also discuss a real-life data example from Sutradhar and Rao (2003) where data are influenced by two random effects due to two factors in a factorial design setup.

5.3.1 Models in Two-Way Factorial Design Setup and Basic Properties

Let y_{ij} denote the response due to the ith $(i = 1,\ldots,m)$ level of a factor A, and the jth $(j = 1,\ldots,n)$ level of a factor B, say, and x_{ij} be a $p \times 1$ vector of covariates associated with y_{ij}. Suppose that conditional on the random variables γ_i and α_j, Y_{ij} follows the binary distribution given by

$$f(y_{ij}|\gamma_i,\alpha_j) = \left[\{\pi^*_{ij}\}^{y_{ij}}\{1-\pi^*_{ij}\}^{1-y_{ij}}\right], \tag{5.97}$$

where

$$\pi^*_{ij} = Pr[Y_{ij} = 1|\gamma_i,\alpha_j] = \frac{\exp(\eta^*_{ij})}{1+\exp(\eta^*_{ij})}$$

with

$$\eta^*_{ij}(\beta,\gamma_i,\alpha_j) = x'_{ij}\beta + \gamma_i + \alpha_j.$$

Further suppose that γ_i and α_j are normally distributed; $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma_\gamma^2)$ and $\alpha_j \overset{\text{iid}}{\sim} N(0, \sigma_\alpha^2)$. Also suppose that γ_i and α_j are independent. For

$$\gamma_i^* = \frac{\gamma_i}{\sigma_\gamma} \text{ and } \alpha_j^* = \frac{\alpha_j}{\sigma_\alpha},$$

one may then express the linear predictor as

$$\eta_{ij}^*(\beta, \gamma_i, \alpha_j) = x_{ij}'\beta + \sigma_\gamma \gamma_i^* + \sigma_\alpha \alpha_j^*. \tag{5.98}$$

Note that the models in (5.97) – (5.98) may be treated as an extension to the binary mixed models given by (5.1) – (5.3). However, in the present setup, the levels m and n are finite. For the asymptotic case, we assume that $K = mn \to \infty$. This model (5.97) – (5.98) involving the regression effects β and the two variance components σ_γ^2 and σ_α^2 is referred to as the GLMM (generalized linear mixed models) with two variance components. Here the scientific interest is to obtain consistent as well as efficient estimates of the regression effects β, and the variance components σ_γ^2 and σ_α^2. Note that obtaining the efficient estimates would require the use of correlation structure of the data.

In the present two-way factorial design setup, the observations are correlated in two ways. More specifically, at the ith $(i = 1, \ldots, m)$ level of the factor A (say), Y_{ij} and Y_{ik} are independent conditional on γ_i^* but unconditionally they are correlated with correlation $\rho_{(ii)jk}$, say. Similarly, at the jth $j(j = 1, \ldots, n)$ level of the factor B (say), Y_{ij} and Y_{rj} are independent conditional on α_j^* but unconditionally they are correlated with correlation $\tilde{\rho}_{ir(jj)}$, say. In general, these correlations $\rho_{(ii)jk}$ and $\tilde{\rho}_{ir(jj)}$ are different. They will be the same in a very special case only when $x_{ij}'\beta = \mu$ (say) for all i and j, and also $\sigma_\gamma^2 = \sigma_\alpha^2$. In Section 5.3.1.2, we provide the formulas for the two-way covariances (or correlations) in the general case. The formulas for the unconditional means and variances are given in Section 5.3.1.1 below. The estimation of the parameters is discussed in Section 5.3.2.

5.3.1.1 Unconditional Mean

Following Lemma 5.1, the unconditional mean and the variance may be written as

$$E[Y_{ij}] = \pi_{ij}(\beta, \sigma_\gamma^2, \sigma_\alpha^2) = \int\int \pi_{ij}^*(\gamma_i^*, \alpha_j^*) g_N(\gamma_i^*|1) \times g_N(\alpha_j^*|1) d\gamma_i^* d\alpha_j^* \tag{5.99}$$

$$\text{var}[Y_{ij}] = \sigma_{(ii)(jj)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2) = \pi_{ij}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)(1 - \pi_{ij}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)), \tag{5.100}$$

where

$$\pi_{ij}^*(\gamma_i^*, \alpha_j^*) = \frac{\exp(\eta_{ij}^*)}{1 + \exp(\eta_{ij}^*)}$$

with

$$\eta_{ij}^* = x_{ij}'\beta + \sigma_\gamma \gamma_i^* + \sigma_\alpha \alpha_i^*,$$

as in (5.97). Note that in the manner similar to that of (5.20), one may evaluate $\pi_{ij}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)$ and $\sigma_{(ii)(jj)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)$, as

$$\pi_{ij}^{(s)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2) = \frac{1}{N} \sum_{w=1}^{N} [\pi_{ij}^*(\gamma_{iw}^*, \alpha_{jw}^*)]$$

$$\sigma_{(ii)(jj)}^{(s)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2) = \pi_{ij}^{(s)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)[1 - \pi_{ij}^{(s)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)], \quad (5.101)$$

where for $w = 1, \ldots, N$, γ_{iw}^* and α_{jw}^*, are two sets of values from the same standard normal distribution.

5.3.1.2 Unconditional Covariances and Correlations in a Two-Way Design Setup

At a given level i of the factor A, the covariance between two responses at the jth and kth levels of the factor B may be written as

$$\text{cov}(Y_{ij}, Y_{ik}) = \lambda_{(ii)jk} - \pi_{ij}\pi_{ik} = \sigma_{(ii)jk} \text{ (say)}, \quad (5.102)$$

where $\lambda_{(ii)jk}$ is given by

$$\lambda_{(ii)jk} = E(Y_{ij}Y_{ik}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\pi_{ij}^*(\gamma_i^*, \alpha_j^*)\pi_{ik}^*(\gamma_i^*, \alpha_k^*)]$$

$$\times g_N(\gamma_i^*|1) g_N(\alpha_j^*|1) g_N(\alpha_k^*|1) d\gamma_i^* d\alpha_j^* d\alpha_k^*, \quad (5.103)$$

which may be computed by using its simulation version

$$\lambda_{(ii)jk}^{(s)} = N^{-1} \sum_{w=1}^{N} [\pi_{ij}^*(\gamma_{iw}^*, \alpha_{jw}^*)\pi_{ik}^*(\gamma_{iw}^*, \alpha_{kw}^*)], \quad (5.104)$$

generating three sets of standardized normal values γ_{iw}^*, α_{jw}^*, and α_{kw}^*, for $w = 1, \ldots, N$.

Note that we have used a slightly different notation for the raw pairwise product moment, namely, $\lambda_{(ii)jk}$ instead of simply λ_{ijk} in (5.9). This is because, in the present two-way design setup, $E[Y_{ij}Y_{ik}]$ for a given i and the $E[Y_{ij}Y_{rj}]$ for a given j do not have the same interpretation. Thus, at a given level j of the factor B, we write the covariance between two responses at the ith and rth levels of the factor A, as

$$\text{cov}(Y_{ij}, Y_{rj}) = E(Y_{ij}Y_{rj}) - \pi_{ij}\pi_{rj} = \lambda_{ir(jj)} - \pi_{ij}\pi_{rj} = \sigma_{ir(jj)} \text{ (say)}, \quad (5.105)$$

where $\lambda_{ir(jj)}$ may be calculated by using the simulated version of $\lambda_{ir(jj)}$ given by

$$\lambda_{ir(jj)}^{(s)} = N^{-1} \sum_{w=1}^{N} [\pi_{ij}^{*}(\gamma_{iw}^{*}, \alpha_{jw}^{*})\pi_{rj}^{*}(\gamma_{rw}^{*}, \alpha_{jw}^{*})], \tag{5.106}$$

where, unlike (5.104), a different set of three standardized normal sequences γ_{iw}^{*}, γ_{rw}^{*}, and α_{jw}^{*}, for $w = 1, \ldots, N$, are used.

Once the covariances are computed by (5.102) and (5.105), the respective correlations may easily be computed as

$$\begin{aligned} \text{corr}[Y_{ij}, Y_{ik}] &= \frac{\sigma_{(ii)jk}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)}{[\pi_{ij}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)(1-\pi_{ij}(\beta, \sigma_\gamma^2))\pi_{ik}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)(1-\pi_{ik}(\beta, \sigma_\gamma^2, \sigma_\alpha^2))]^{1/2}} \\ &= \rho_{(ii)jk} \text{ (say).} \end{aligned} \tag{5.107}$$

and

$$\begin{aligned} \text{corr}[Y_{ij}, Y_{rj}] &= \frac{\sigma_{ir(jj)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)}{[\pi_{ij}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)(1-\pi_{ij}(\beta, \sigma_\gamma^2))\pi_{rj}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)(1-\pi_{rj}(\beta, \sigma_\gamma^2, \sigma_\alpha^2))]^{1/2}} \\ &= \rho_{ir(jj)} \text{ (say).} \end{aligned} \tag{5.108}$$

5.3.2 Estimation of Parameters

5.3.2.1 Estimation of Regression Effects β

Let $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{ik}, \ldots, y_{in})'$ be the n-dimensional response vector at the ith $(i = 1, \ldots, m)$ level of factor A and $\pi_i = (\pi_{i1}, \ldots, \pi_{ij}, \ldots, \pi_{ik}, \ldots, \pi_{in})'$ be the corresponding unconditional mean vector. Use y_i and write a stack vector as

$$y = (y_1', \ldots, y_i', \ldots, y_r', \ldots, y_m')'$$

of dimension $mn \times 1$ which has its mean vector

$$\pi = (\pi_1', \ldots, \pi_i', \ldots, \pi_r', \ldots, \pi_m')'.$$

Define the $mn \times mn$ covariance matrix of Y as

$$\Sigma = (\text{cov}[Y_{ij}, Y_{rk}]) = \begin{cases} \sigma_{(ii)(jj)} & \text{for } i = r; j = k \\ \sigma_{(ii)jk} & \text{for } i = r; j \neq k \\ \sigma_{ir(jj)} & \text{for } i \neq r; j = k \\ 0 & \text{for } i \neq r; j \neq k, \end{cases} \tag{5.109}$$

where the formulas for the variances $\sigma_{(ii)(jj)}$, and covariances $\sigma_{(ii)jk}$ and $\sigma_{ir(jj)}$, are given by (5.100), (5.102), and (5.105), respectively.

In order to obtain a consistent as well as efficient estimate for β, we take the two-way covariances given by (5.109) into account and construct the GQL estimating equation as

$$D'\Sigma^{-1}(y-\pi) = 0, \tag{5.110}$$

where $D = \partial\pi/\partial\beta'$ is the $mn \times p$ first derivative matrix of the stacked mean vector π with respect to β. Note that the formulas for the elements of the derivative matrix can be computed by using the general formula, say for the derivative of $\pi_{ij}(\beta, \sigma^2_\gamma, \sigma^2_\alpha)$ with respect to β. This formula is given by

$$\begin{aligned}\frac{\partial\pi_{ij}(\beta, \sigma^2_\gamma, \sigma^2_\alpha)}{\partial\beta} &= \int\int \pi^*_{ij}(\gamma^*_i, \alpha^*_j)[1 - \pi^*_{ij}(\gamma^*_i, \alpha^*_j)]x_{ij} \\ &\quad \times g_N(\gamma^*_i|1)g_N(\alpha^*_j|1)d\gamma^*_i d\alpha^*_j.\end{aligned} \tag{5.111}$$

As indicated before, we now replace the π vector, Σ, and derivative (D) matrices in (5.110) with their simulated versions $\pi^{(s)}$, $\Sigma^{(s)}$, and $D^{(s)}$, respectively, and write the simulated version of the GQL estimating equation as

$$D^{(s)'}\{\Sigma^{(s)}\}^{-1}(y - \pi^{(s)}) = 0, \tag{5.112}$$

which may be solved iteratively by using the Newton−Raphson procedure. Note that the construction of $\pi^{(s)}$ follows by using the $\pi^{(s)}_{ij}(\beta, \sigma^2_\gamma, \sigma^2_\alpha)$ from (5.101). Next, $\Sigma^{(s)}$ can be computed by using the formulas for $\pi^{(s)}_{ij}(\beta, \sigma^2_\gamma, \sigma^2_\alpha)$ from (5.101), and the formulas for $\lambda^{(s)}_{(ii)jk}$ and $\lambda^{(s)}_{ir(jj)}$, from (5.104) and (5.106), respectively. Similarly, $D^{(s)}$ is computed by using the simulated version of the formula (5.111) given by

$$\frac{\partial\pi^{(s)}_{ij}(\beta, \sigma^2_\gamma, \sigma^2_\alpha)}{\partial\beta} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})[1 - \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})]x_{ij}. \tag{5.113}$$

Suppose that $\hat{\beta}_{GQL}$ denotes the solution of the GQL estimating equation (5.112). As the GQL estimating equation (5.112) is unbiased for the estimation of β, this solution $\hat{\beta}_{GQL}$ is consistent for β. Furthermore, as the estimating equation (5.112) for β is constructed by taking the two-way correlations of the data into account, $\hat{\beta}_{GQL}$ is highly efficient too, the exact maximum likelihood estimator being the most efficient. The exact maximum likelihood estimation is, however, more complicated than the present GQL estimation. Furthermore, for $K = mn$, it may be shown under some mild regularity conditions that $K^{1/2}(\hat{\beta}_{GQL} - \beta)$ has an asymptotic normal distribution, as $K \to \infty$, with mean zero and a covariance matrix

$$G_1 = K\left[D'\Sigma^{-1}D\right]^{-1},$$

that can be computed by

$$G_1^{(s)} = K\left[D^{(s)'}\{\Sigma^{(s)}\}^{-1}D^{(s)}\right]^{-1}. \qquad (5.114)$$

5.3.2.2 Estimation of the Variance Component σ_γ^2 Due to Factor A

For $i = 1,\ldots,m$, let $u_{i(s)} = (y_{i1}^2,\ldots,y_{ij}^2,\ldots,y_{in}^2)'$ be the n-dimensional vector of squares of the elements of y_i and $u_{i(p)} = (y_{i1}y_{i2},\ldots,y_{ij}y_{ik},\ldots,y_{i,n-1}y_{in})'$ be the $n(n-1)/2$-dimensional vector of cross-products for all n responses under the ith level of factor A. Furthermore, let $s_i = (u'_{i(s)}, u'_{i(p)})'$ be the $n(n+1)/2$-dimensional combined vector of squares and products. Note that because $u_{i(s)} = (y_{i1}^2,\ldots,y_{ij}^2,\ldots,y_{in}^2)' \equiv y_i = (y_{i1},\ldots,y_{ij},\ldots,y_{in})'$ for the binary data, we use the basic statistic s_i given by

$$s_i = [y_i', u'_{i(p)}]' \qquad (5.115)$$

and its properties to develop the desired GQL estimating equation for σ_γ^2.

Let λ_i^* denote the expectation of s_i. Because $E[Y_{ij}] = \pi_{ij}(\beta,\sigma_\gamma^2,\sigma_\alpha^2)$ as in (5.99) and $E[Y_{ij}Y_{ik}] = \lambda_{(ii)jk}(\beta,\sigma_\gamma^2,\sigma_\alpha^2)$ by (5.102), we can write

$$E[S_i] = \lambda_i^* = [\pi_i', \lambda'_{(ii)}]' :\ n(n+1)/2 \times 1, \qquad (5.116)$$

where

$$\begin{aligned}\pi_i &= [\pi_{i1},\ldots,\pi_{ij},\ldots,\pi_{in}]' \\ \lambda_{(ii)} &= [\lambda_{(ii)12},\ldots,\lambda_{(ii)jk},\ldots,\lambda_{(ii)(n-1)n}]'. \end{aligned} \qquad (5.117)$$

In addition, we construct a stacked vector

$$s = (s_1',\ldots,s_i',\ldots,s_r',\ldots,s_m')' :\ mn(n+1)/2 \times 1.$$

Let

$$\begin{aligned}\lambda^* &= E[S] = [\lambda^{*\prime}_1,\ldots,\lambda^{*\prime}_i,\ldots,\lambda^{*\prime}_r,\ldots,\lambda^{*\prime}_m]' :\ mn(n+1)/2 \times 1 \qquad (5.118)\\ \Omega^* &= \mathrm{cov}[S] \\ &= \begin{bmatrix} \mathrm{cov}(S_1) & \ldots & \mathrm{cov}(S_1,S_i') & \ldots & \mathrm{cov}(S_1,S_r') & \ldots & \mathrm{cov}(S_1,S_m') \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{cov}(S_i,S_1') & \ldots & \mathrm{cov}(S_i) & \ldots & \mathrm{cov}(S_i,S_r') & \ldots & \mathrm{cov}(S_i,S_m') \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{cov}(S_m,S_1') & \ldots & \mathrm{cov}(S_m,S_i') & \ldots & \mathrm{cov}(S_m,S_r') & \ldots & \mathrm{cov}(S_m) \end{bmatrix}\end{aligned}$$

$$= \begin{bmatrix} \Omega^*_{11} & \ldots & \Omega^*_{1i} & \ldots & \Omega^*_{1r} & \ldots & \Omega^*_{1m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \Omega^*_{i1} & \ldots & \Omega^*_{ii} & \ldots & \Omega^*_{ir} & \ldots & \Omega^*_{im} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \Omega^*_{m1} & \ldots & \Omega^*_{mi} & \ldots & \Omega^*_{mr} & \ldots & \Omega^*_{mm} \end{bmatrix}, \tag{5.119}$$

where

$$\Omega^*_{ii} = \begin{bmatrix} \mathrm{cov}(Y_i) & \mathrm{cov}(Y_i, U'_{i(p)}) \\ & \mathrm{cov}(U_{i(p)}) \end{bmatrix} = \begin{bmatrix} \Sigma^*_{ii,11} & \Sigma^*_{ii,12} \\ & \Sigma^*_{ii,22} \end{bmatrix} \text{ (say)}, \tag{5.120}$$

and for $i \neq r$,

$$\Omega^*_{ir} = \begin{bmatrix} \mathrm{cov}(Y_i, Y'_r) & \mathrm{cov}(Y_i, U'_{r(p)}) \\ & \mathrm{cov}(U_{r(p)}) \end{bmatrix} = \begin{bmatrix} \Sigma^*_{ir,11} & \Sigma^*_{ir,12} \\ & \Sigma^*_{ir,22} \end{bmatrix} \text{ (say)}. \tag{5.121}$$

Next, let $B^* = \partial\lambda^*/\partial\sigma^2_\gamma$ be the $((mn(n+1))/2) \times 1$ vector of first derivatives of the elements of λ^*. It then follows that the GQL estimating equation for σ^2_γ is given by

$$B^{*\prime}\Omega^{*-1}(s - \lambda^*) = 0. \tag{5.122}$$

Note, however, that as the components of the λ^* vector and Ω^* and B^* matrices involve integrations over the distributions of the random effects, which are not easy to evaluate, we use their simulated approximations, namely $\lambda^{*(s)}$, $\Omega^{*(s)}$, and $B^{*(s)}$, and rewrite the GQL estimating equation (5.122) for σ^2_γ as

$$B^{*(s)\prime}\{\Omega^{*(s)}\}^{-1}(s - \lambda^{*(s)}) = 0. \tag{5.123}$$

Suppose that $\hat{\sigma}^2_{\gamma,GQL}$ is the solution of the estimating equation (5.123) for σ^2_γ. By arguments similar to those for $\hat{\beta}_{GQL}$ obtained from (5.112), this solution $\hat{\sigma}^2_{\gamma,GQL}$ obtained from the GQL estimating equation (5.123) is a consistent estimator for σ^2_γ, and it is also efficient. Note that we still need to compute $\Omega^{*(s)}$ for (5.123), which is given below. Furthermore, the formulas for the elements of the $B^* = \partial\lambda^*/\partial\sigma^2_\gamma$

matrix are available in Exercise 5.3.

Construction of $\Omega^{*(s)}$:
The construction of the $\Omega^{*(s)}$ matrix requires the formulas for the $\Omega^{*(s)}_{ii}$ (5.120) and $\Omega^{*(s)}_{ir}$ matrices. We first provide the formulas for the component matrices for $\Omega^{*(s)}_{ii}$.

Formulas for the elements of $\Omega^{*(s)}_{ii}$
Construction of $\Sigma^{*(s)}_{ii,11}$ $: n \times n$

$$\Sigma^{*(s)}_{ii,11} = (\text{cov}[Y_{ij}, Y_{ik}]) = \begin{cases} \sigma^{(s)}_{(ii)(jj)} = \pi^{(s)}_{ij}[1 - \pi^{(s)}_{ij}] & \text{for } j = k \\ \sigma^{(s)}_{(ii)jk} = \lambda^{(s)}_{(ii)jk} - \pi^{(s)}_{ij}\pi^{(s)}_{ik} & \text{for } j \neq k, \end{cases} \tag{5.124}$$

where $\pi^{(s)}_{ij}$ and $\lambda^{(s)}_{(ii)jk}$ are given in (5.101) and (5.104), respectively.

Construction of $\Sigma^{*(s)}_{ii,12}$ $: n \times n(n-1)/2$

$$\Sigma^{*(s)}_{ii,12} = (\text{cov}[Y_{ij}, Y_{ik}Y_{i\ell}]) = \begin{cases} \lambda^{(s)}_{(ii)j\ell} - \pi^{(s)}_{ij}[\lambda^{(s)}_{(ii)j\ell}] & \text{for } j = k \\ \lambda^{(s)}_{(ii)jk} - \pi^{(s)}_{ij}[\lambda^{(s)}_{(ii)jk}] & \text{for } j = \ell \\ \delta^{(s)}_{(ii)jk\ell} - \pi^{(s)}_{ij}[\lambda^{(s)}_{(ii)k\ell}] & \text{for } j \neq k; j \neq \ell, \end{cases} \tag{5.125}$$

where

$$\delta^{(s)}_{(ii)jk\ell} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{ik}(\gamma^*_{iw}, \alpha^*_{kw})\pi^*_{i\ell}(\gamma^*_{iw}, \alpha^*_{\ell w}).$$

Construction of $\Sigma^{*(s)}_{ii,22}$ $: n(n-1)/2 \times n(n-1)/2$

$$\Sigma^{*(s)}_{ii,22} = (\text{cov}[Y_{ij}Y_{ik}, Y_{i\ell}Y_{iv}]) = \begin{cases} \lambda^{(s)}_{(ii)jk} - [\lambda^{(s)}_{(ii)jk}]^2 & \text{for } j = \ell; k = v \\ \delta^{(s)}_{(ii)jkv} - [\lambda^{(s)}_{(ii)jk}][\lambda^{(s)}_{(ii)jv}] & \text{for } j = \ell; k \neq v \\ \delta^{(s)}_{(ii)jk\ell} - [\lambda^{(s)}_{(ii)jk}][\lambda^{(s)}_{(ii)j\ell}] & \text{for } j = v; k \neq \ell \\ \phi^{(s)}_{(ii)jk\ell v} - [\lambda^{(s)}_{(ii)jk}][\lambda^{(s)}_{(ii)\ell v}] & \text{for } j \neq \ell; k \neq v, \end{cases} \tag{5.126}$$

where

$$\phi^{(s)}_{(ii)jk\ell v} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{ik}(\gamma^*_{iw}, \alpha^*_{kw})\pi^*_{i\ell}(\gamma^*_{iw}, \alpha^*_{\ell w})\pi^*_{iv}(\gamma^*_{iw}, \alpha^*_{vw}).$$

Formulas for the elements of $\Omega^{*(s)}_{ir}$ for $i \neq r$
Construction of $\Sigma^{*(s)}_{ir,11}$ $: n \times n$

$$\Sigma^{*(s)}_{ir,11} = (\text{cov}[Y_{ij}, Y_{rk}]) = \begin{cases} \sigma^{(s)}_{ir(jj)} = \lambda^{(s)}_{ir(jj)} - \pi^{(s)}_{ij}\pi^{(s)}_{rj} & \text{for } j = k \\ \sigma^{(s)}_{(ir)jk} = 0 & \text{for } j \neq k, \end{cases} \tag{5.127}$$

where $\pi^{(s)}_{ij}$ and $\lambda^{(s)}_{ir(jj)}$ are given in (5.101) and (5.106), respectively.

Construction of $\Sigma^{*(s)}_{ir,12} : n \times n(n-1)/2$

$$\Sigma^{*(s)}_{ir,12} = (\text{cov}[Y_{ij}, Y_{rk}Y_{r\ell}]) = \begin{cases} \delta^{(s)}_{(irr)jj\ell} - \pi^{(s)}_{ij}[\lambda^{(s)}_{(rr)j\ell}] & \text{for } j = k \\ \delta^{(s)}_{(irr)jkj} - \pi^{(s)}_{ij}[\lambda^{(s)}_{(rr)jk}] & \text{for } j = \ell \\ 0 & \text{for } j \neq k; j \neq \ell, \end{cases} \tag{5.128}$$

where

$$\delta^{(s)}_{(irr)jj\ell} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{rj}(\gamma^*_{rw}, \alpha^*_{jw})\pi^*_{r\ell}(\gamma^*_{rw}, \alpha^*_{\ell w}),$$

and

$$\delta^{(s)}_{(irr)jkj} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{rk}(\gamma^*_{rw}, \alpha^*_{kw})\pi^*_{rj}(\gamma^*_{rw}, \alpha^*_{jw}).$$

Construction of $\Sigma^{*(s)}_{ir,22} : n(n-1)/2 \times n(n-1)/2$

$$\Sigma^{*(s)}_{ir,22} = (\text{cov}[Y_{ij}Y_{ik}, Y_{r\ell}Y_{rv}]) = \begin{cases} \phi^{(s)}_{(iirr)jkjk} - [\lambda^{(s)}_{(ii)jk}\lambda^{(s)}_{(rr)jk}] & \text{for } j = \ell; k = v \\ \phi^{(s)}_{(iirr)jkjv} - [\lambda^{(s)}_{(ii)jk}][\lambda^{(s)}_{(rr)jv}] & \text{for } j = \ell; k \neq v \\ \phi^{(s)}_{(iirr)jkj\ell} - [\lambda^{(s)}_{(ii)jk}][\lambda^{(s)}_{(rr)j\ell}] & \text{for } j = v; k \neq \ell \\ 0 & \text{for } j \neq \ell; k \neq v, \end{cases} \tag{5.129}$$

where, for example,

$$\phi^{(s)}_{(iirr)jkjk} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{ik}(\gamma^*_{iw}, \alpha^*_{kw})\pi^*_{rj}(\gamma^*_{rw}, \alpha^*_{jw})\pi^*_{rk}(\gamma^*_{rw}, \alpha^*_{kw}).$$

We now turn back to the properties of $\hat{\sigma}^2_{\gamma,GQL}$ obtained from (5.123). For $K = mn$, under mild regularity conditions, it may be shown that as $K \to \infty$, $K^{1/2}(\hat{\sigma}^2_{\gamma,GQL} - \sigma^2_\gamma)$ has asymptotically a univariate normal distribution with mean zero and the variance that may be computed by

$$G^{(s)}_2 = K[\{B^{*(s)'}\}\{\Omega^{*(s)}\}^{-1}\{B^{*(s)}\}]^{-1}. \tag{5.130}$$

5.3.2.3 Estimation of the Variance Component σ_α^2 Due to Factor B

Note that the variance component of the factor B is usually referred to as the variance component of the column effects. To estimate this variance component, as mentioned earlier, we now exploit the combined vector of squares and pairwise products of the observations recorded under the columns, whereas σ_γ^2 was computed in Section 5.3.2.2 by utilizing responses recorded under the rows. To be specific, use the responses of the jth $(j = 1,\ldots,n)$ column and define

$$\tilde{s}_j = (\tilde{u}'_{j(s)}, \tilde{u}'_{j(p)})', \tag{5.131}$$

where

$$\tilde{u}_{j(s)} = (y_{1j}^2,\ldots,y_{ij}^2,\ldots,y_{mj}^2)'$$

and

$$\tilde{u}_{j(p)} = (y_{1j}y_{2j},\ldots,y_{ij}y_{rj},\ldots,y_{m-1,j}y_{mj})'.$$

Note that for the present binary case, the vector statistic in (5.131) is equivalent to

$$\tilde{s}_j = (\tilde{y}'_j, \tilde{u}'_{j(p)})', \tag{5.132}$$

where $\tilde{y}_j = [y_{1j},\ldots,y_{ij},\ldots,y_{mj}]'$.

Let $\tilde{\lambda}_i$ denote the expectation of $\tilde{s}_j$. Because $E[Y_{ij}] = \pi_{ij}(\beta,\sigma_\gamma^2,\sigma_\alpha^2)$ as in (5.99) and $E[Y_{ij}Y_{rj}] = \lambda_{ir(jj)}(\beta,\sigma_\gamma^2,\sigma_\alpha^2)$ by (5.105), we can write

$$E[\tilde{S}_j] = \tilde{\lambda}_j = [\tilde{\pi}'_j, \tilde{\lambda}'_{(jj)}]' : \; m(m+1)/2 \times 1, \tag{5.133}$$

where

$$\begin{aligned} \tilde{\pi}_j &= [\pi_{1j},\ldots,\pi_{ij},\ldots,\pi_{mj}]' \\ \tilde{\lambda}_{(jj)} &= [\lambda_{12(jj)},\ldots,\lambda_{ir(jj)},\ldots,\lambda_{(m-1)m(jj)}]'. \end{aligned} \tag{5.134}$$

Furthermore, we construct a stacked vector

$$\tilde{s} = (\tilde{s}'_1,\ldots,\tilde{s}'_i,\ldots,\tilde{s}'_r,\ldots,\tilde{s}'_n)' : \; nm(m+1)/2 \times 1.$$

Let

$$\begin{aligned} \tilde{\lambda} &= E[\tilde{S}] = [\tilde{\lambda}'_1,\ldots,\tilde{\lambda}'_i,\ldots,\tilde{\lambda}'_r,\ldots,\tilde{\lambda}'_n]' : \; nm(m+1)/2 \times 1 \qquad (5.135) \\ \tilde{\Omega} &= \mathrm{cov}[\tilde{S}] \end{aligned}$$

$$= \begin{bmatrix} \mathrm{cov}(\tilde{S}_1) & \ldots & \mathrm{cov}(\tilde{S}_1,\tilde{S}'_j) & \ldots & \mathrm{cov}(\tilde{S}_1,\tilde{S}'_k) & \ldots & \mathrm{cov}(\tilde{S}_1,\tilde{S}'_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\tilde{S}_j,\tilde{S}'_1) & \ldots & \mathrm{cov}(\tilde{S}_j) & \ldots & \mathrm{cov}(\tilde{S}_j,\tilde{S}'_k) & \ldots & \mathrm{cov}(\tilde{S}_j,\tilde{S}'_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\tilde{S}_n,\tilde{S}'_1) & \ldots & \mathrm{cov}(\tilde{S}_n,\tilde{S}'_j) & \ldots & \mathrm{cov}(\tilde{S}_n,\tilde{S}'_k) & \ldots & \mathrm{cov}(\tilde{S}_n) \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{\Omega}_{11} & \ldots & \tilde{\Omega}_{1j} & \ldots & \tilde{\Omega}_{1k} & \ldots & \tilde{\Omega}_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\Omega}_{j1} & \ldots & \tilde{\Omega}_{jj} & \ldots & \tilde{\Omega}_{jk} & \ldots & \tilde{\Omega}_{jn} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\Omega}_{n1} & \ldots & \tilde{\Omega}_{nj} & \ldots & \tilde{\Omega}_{nk} & \ldots & \tilde{\Omega}_{nn} \end{bmatrix}, \tag{5.136}$$

where

$$\tilde{\Omega}_{jj} = \begin{bmatrix} \mathrm{cov}(\tilde{Y}_j) & \mathrm{cov}(\tilde{Y}_j,\tilde{U}'_{j(p)}) \\ & \mathrm{cov}(\tilde{U}_{j(p)}) \end{bmatrix} = \begin{bmatrix} \tilde{\Sigma}_{jj,11} & \tilde{\Sigma}_{jj,12} \\ & \tilde{\Sigma}_{jj,22} \end{bmatrix} \text{ (say)}, \tag{5.137}$$

and for $j \neq k$,

$$\tilde{\Omega}_{jk} = \begin{bmatrix} \mathrm{cov}(\tilde{Y}_j,\tilde{Y}'_k) & \mathrm{cov}(\tilde{Y}_j,\tilde{U}'_{k(p)}) \\ & \mathrm{cov}(\tilde{U}_{k(p)}) \end{bmatrix} = \begin{bmatrix} \tilde{\Sigma}_{jk,11} & \tilde{\Sigma}_{jk,12} \\ & \tilde{\Sigma}_{jk,22} \end{bmatrix} \text{ (say)}. \tag{5.138}$$

Let $\tilde{B} = \partial\tilde{\lambda}/\partial\sigma^2_\alpha$ be the $((nm(m+1))/2) \times 1$ vector of first derivatives of the elements of $\tilde{\lambda}$. It then follows that the GQL estimating equation for σ^2_α is given by

$$\tilde{B}'\tilde{\Omega}^{-1}(\tilde{s}-\tilde{\lambda}) = 0. \tag{5.139}$$

In the manner similar to that of (5.123), we may now rewrite the GQL estimating equation (5.139) for σ^2_α as

$$\tilde{B}^{(s)'}\{\tilde{\Omega}^{(s)}\}^{-1}(\tilde{s}-\tilde{\lambda}^{(s)}) = 0. \tag{5.140}$$

Suppose that $\hat{\sigma}^2_{\alpha,GQL}$ is the solution of the estimating equation (5.140) for σ^2_α. By arguments similar to those for $\hat{\beta}_{GQL}$ obtained from (5.112), this solution $\hat{\sigma}^2_{\alpha,GQL}$ obtained from the GQL estimating equation (5.140) is a consistent estimator for σ^2_α, and it is also efficient.

As far as the asymptotic distribution of $\hat{\sigma}^2_{\alpha,GQL}$ is concerned, one may obtain this in a manner similar to that of $\hat{\sigma}^2_{\gamma,GQL}$. More specifically, similar to (5.130), for $K = mn$, under mild regularity conditions, it may be shown that as $K \rightarrow \infty$, $K^{\frac{1}{2}}(\hat{\sigma}^2_{\alpha,GQL} - \sigma^2_\alpha)$ has asymptotically a univariate normal distribution with mean zero and the variance that may be computed by

$$G_3^{(s)} = K[\{\tilde{B}^{(s)'}\}\{\tilde{\Omega}^{(s)}\}^{-1}\{\tilde{B}^{(s)}\}]^{-1}. \quad (5.141)$$

Note that the estimating equation (5.140) and the asymptotic distribution of $\hat{\sigma}^2_{\alpha,GQL}$ in (5.141) still require the formulas for $\tilde{\Omega}^{(s)}$ and $\tilde{B} = \partial\tilde{\lambda}/\partial\sigma^2_\alpha$. The formulas for the elements of the derivative matrix $\tilde{B}^{(s)}$ are available from Exercise 5.4, whereas we provide the formulas for the elements of $\tilde{\Omega}^{(s)}$ matrix as follows.

Construction of $\tilde{\Omega}^{(s)}$

The construction of the $\tilde{\Omega}^{(s)}$ matrix requires the formulas for the $\tilde{\Omega}^{(s)}_{jj}$ (5.137) and $\tilde{\Omega}^{(s)}_{jk}$ (5.138) matrices. We first provide the formulas for the component matrices for $\tilde{\Omega}^{(s)}_{jj}$.

Formulas for the elements of $\tilde{\Omega}^{(s)}_{jj}$

Construction of $\tilde{\Sigma}^{(s)}_{jj,11}$: $m \times m$

$$\tilde{\Sigma}^{(s)}_{jj,11} = (\mathrm{cov}[Y_{ij}, Y_{rj}]) = \begin{cases} \sigma^{(s)}_{(ii)(jj)} = \pi^{(s)}_{ij}[1 - \pi^{(s)}_{ij}] & \text{for } i = r \\ \sigma^{(s)}_{ir(jj)} = \lambda^{(s)}_{ir(jj)} - \pi^{(s)}_{ij}\pi^{(s)}_{rj} & \text{for } i \neq r, \end{cases} \quad (5.142)$$

where $\pi^{(s)}_{ij}$ and $\lambda^{(s)}_{ir(jj)}$ are given in (5.101) and (5.106), respectively.

Construction of $\tilde{\Sigma}^{(s)}_{jj,12}$: $n \times m(m-1)/2$

$$\tilde{\Sigma}^{(s)}_{jj,12} = (\mathrm{cov}[Y_{ij}, Y_{rj}Y_{uj}]) = \begin{cases} \lambda^{(s)}_{iu(jj)} - \pi^{(s)}_{ij}[\lambda^{(s)}_{iu(jj)}] & \text{for } r = i \\ \lambda^{(s)}_{ir(jj)} - \pi^{(s)}_{ij}[\lambda^{(s)}_{ir(jj)}] & \text{for } u = i \\ \delta^{(s)}_{iru(jj)} - \pi^{(s)}_{ij}[\lambda^{(s)}_{ru(jj)}] & \text{for } r \neq i; u \neq i, \end{cases} \quad (5.143)$$

where

$$\delta^{(s)}_{iru(jj)} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{rj}(\gamma^*_{rw}, \alpha^*_{jw})\pi^*_{uj}(\gamma^*_{uw}, \alpha^*_{jw}).$$

Construction of $\tilde{\Sigma}^{(s)}_{jj,22} : m(m-1)/2 \times m(m-1)/2$

$$\tilde{\Sigma}^{(s)}_{jj,22} = (\mathrm{cov}[Y_{ij}Y_{rj}, Y_{uj}Y_{zj}]) = \begin{cases} \lambda^{(s)}_{ir(jj)} - [\lambda^{(s)}_{ir(jj)}]^2 & \text{for } i = u; r = z \\ \delta^{(s)}_{irz(jj)} - [\lambda^{(s)}_{ir(jj)}][\lambda^{(s)}_{iz(jj)}] & \text{for } i = u; r \neq z \\ \delta^{(s)}_{iru(jj)} - [\lambda^{(s)}_{ir(jj)}][\lambda^{(s)}_{iu(jj)}] & \text{for } i = z; r \neq u \\ \phi^{(s)}_{iruz(jj)} - [\lambda^{(s)}_{ir(jj)}][\lambda^{(s)}_{uz(jj)}] & \text{for } i \neq u; r \neq z, \end{cases} \tag{5.144}$$

where

$$\phi^{(s)}_{iruz(jj)} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{rj}(\gamma^*_{rw}, \alpha^*_{jw})\pi^*_{uj}(\gamma^*_{uw}, \alpha^*_{jw})\pi^*_{zj}(\gamma^*_{zw}, \alpha^*_{jw}).$$

Formulas for the Elements of $\tilde{\Omega}^{(s)}_{jk}$ for $j \neq k$

Construction of $\tilde{\Sigma}^{(s)}_{jk,11} : m \times m$

$$\tilde{\Sigma}^{(s)}_{jk,11} = (\mathrm{cov}[Y_{ij}, Y_{rk}]) = \begin{cases} \sigma^{(s)}_{ii(jk)} = \lambda^{(s)}_{ii(jk)} - \pi^{(s)}_{ij}\pi^{(s)}_{ik} & \text{for } i = r \\ \sigma^{(s)}_{ir(jk)} = 0 & \text{for } i \neq r, \end{cases} \tag{5.145}$$

where $\pi^{(s)}_{ij}$ and $\lambda^{(s)}_{ii(jk)} \equiv \lambda^{(s)}_{(ii)jk}$ are given in (5.101) and (5.104), respectively.

Construction of $\tilde{\Sigma}^{(s)}_{jk,12} : m \times m(m-1)/2$

$$\tilde{\Sigma}^{(s)}_{jk,12} = (\mathrm{cov}[Y_{ij}, Y_{rk}Y_{uk}]) = \begin{cases} \delta^{(s)}_{iiu(jkk)} - \pi^{(s)}_{ij}[\lambda^{(s)}_{iu(kk)}] & \text{for } r = i \\ \delta^{(s)}_{iir(jkk)} - \pi^{(s)}_{ij}[\lambda^{(s)}_{ir(kk)}] & \text{for } u = i \\ 0 & \text{for } i \neq r; i \neq u, \end{cases} \tag{5.146}$$

where

$$\delta^{(s)}_{iiu(jkk)} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{ik}(\gamma^*_{iw}, \alpha^*_{kw})\pi^*_{uk}(\gamma^*_{uw}, \alpha^*_{kw}),$$

and

$$\delta^{(s)}_{iir(jkk)} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{rk}(\gamma^*_{rw}, \alpha^*_{kw})\pi^*_{ik}(\gamma^*_{iw}, \alpha^*_{kw}).$$

Construction of $\tilde{\Sigma}^{(s)}_{jk,22} : m(m-1)/2 \times m(m-1)/2$

$$\tilde{\Sigma}^{(s)}_{jk,22} = (\text{cov}[Y_{ij}Y_{rj}, Y_{uk}Y_{zk}]) = \begin{cases} \phi^{(s)}_{irir(jjkk)} - [\lambda^{(s)}_{ir(jj)}\lambda^{(s)}_{ir(kk)}] & \text{for } u = i; z = r \\ \phi^{(s)}_{iriz(jjkk)} - [\lambda^{(s)}_{ir(jj)}][\lambda^{(s)}_{iz(kk)}] & \text{for } u = i; z \neq r \\ \phi^{(s)}_{irui(jjkk)} - [\lambda^{(s)}_{ir(jj)}][\lambda^{(s)}_{iu(kk)}] & \text{for } z = i; r \neq u \\ 0 & \text{for } i \neq u; r \neq z, \end{cases} \tag{5.147}$$

where, for example,

$$\phi^{(s)}_{irir(jjkk)} = \frac{1}{N}\sum_{w=1}^{N} \pi^*_{ij}(\gamma^*_{iw}, \alpha^*_{jw})\pi^*_{rj}(\gamma^*_{rw}, \alpha^*_{jw})\pi^*_{ik}(\gamma^*_{iw}, \alpha^*_{kw})\pi^*_{rk}(\gamma^*_{rw}, \alpha^*_{kw}).$$

5.3.2.4 Computational Steps

Note that the GQL estimators $\hat{\beta}_{GQL}$, $\hat{\sigma}^2_{\gamma,GQL}$, and $\hat{\sigma}^2_{\alpha,GQL}$ for β, σ^2_γ, and σ^2_α are the solutions of the estimating equations (5.112), (5.123), and (5.140), respectively. We obtain these solutions, based on a three-step procedure given below.

Step 1. For suitable initial values of σ^2_γ and σ^2_α, we solve the estimating equation (5.112) for β, by using the iterative equation

$$\hat{\beta}_{GQL}(t+1) = \hat{\beta}_{GQL}(t) + \left[D^{(s)'}\{\Sigma^{(s)}\}^{-1}D^{(s)}\right]^{-1}_t \times \left[D^{(s)'}\{\Sigma^{(s)}\}^{-1}(y-\mu^{(s)})\right]_t, \tag{5.148}$$

where $\hat{\beta}_{GQL}(t)$ denotes the quasi-likelihood estimate of β at the tth iteration, and $[\,]_t$ denotes that the expression within the brackets is evaluated at $\beta = \hat{\beta}_{GQL}(t)$.

Step 2. For the initial value of σ^2_α used in step 1, and for the estimate of β obtained from step 1, we now solve the GQL estimating equation (5.123) for σ^2_γ by using the iterative formula

$$\hat{\sigma}^2_{\gamma,GQL}(t+1) = \hat{\sigma}^2_{\gamma,GQL}(t) + \left[B^{*(s)'}\{\Omega^{*(s)}\}^{-1}\{B^{*(s)}\}\right]^{-1}_t \times \left[B^{*(s)'}\{\Omega^{*(s)}\}^{-1}(s-\lambda^{*(s)})\right]_t, \tag{5.149}$$

where $[\,]_t$ denotes that the expression within the brackets is evaluated at $\sigma^2_\gamma = \hat{\sigma}^2_{\gamma,GQL}(t)$.

Step 3. By using the estimates of β and σ^2_γ obtained from steps 1 and 2, respectively, we solve the GQL estimating equation (5.140) iteratively for σ^2_α, by using the Newton−Raphson iterative formula

$$\hat{\sigma}^2_{\alpha,GQL}(t+1) = \hat{\sigma}^2_{\alpha,GQL}(t) + \left[\tilde{B}^{(s)'}\{\tilde{\Omega}^{(s)}\}^{-1}\{\tilde{B}^{(s)}\}\right]^{-1}_t$$

$$\times \left[\tilde{B}^{(s)'} \{\tilde{\Omega}^{(s)}\}^{-1} (\tilde{s} - \tilde{\lambda}^{(s)}) \right]_t, \quad (5.150)$$

where $[\ ]_t$ denotes that the expression within the brackets is evaluated at $\sigma^2_\alpha = \hat{\sigma}^2_{\alpha,GQL}(t)$.

Next the estimates of σ^2_γ and σ^2_α obtained from steps 2 and 3 are used in step 1 to obtain a new β estimate. This improved β estimate and the estimate of σ^2_α obtained from step 3 are then used in step 2 to obtain an improved estimate of σ^2_γ. Similarly, the improved estimates of β and σ^2_γ are used in step 3 to obtain an improved estimate of σ^2_α. This cycle of iterations continues until convergence of all three estimates. The final solutions are denoted by $\hat{\beta}_{GQL}$, $\hat{\sigma}^2_{\gamma,GQL}$, and $\hat{\sigma}^2_{\alpha,GQL}$ for β, σ^2_γ, and σ^2_α, respectively.

5.3.3 Salamander Mating Data Analysis

5.3.3.1 Data Description

The salamander mating data were recorded from three experiments involving two geographically isolated populations of salamanders, Rough Butt (RB) and White Side (WS). Altogether 10 RB males (RBM) and 10 WS males (WSM) were sequestered as pairs with 10 RB females (RBF) and 10 WS females (WSF) on six occasions according to a design given in McCullagh and Nelder [1989, Table 14.3]. For each pair, it was recorded whether mating occurred. All 40 animals mentioned above were used in each of three experiments, one conducted in the summer of 1986 and two in the fall of the same year, but the animals used in the first fall experiment were identical to those used in the summer experiment. This certainly introduces longitudinal correlations between the binary responses (1 for occurrence of mating and 0 for nonoccurrence) repeatedly collected from each fixed pair of animals over two time points. Kuk (1995), for illustration, analyzed the mating data from the summer experiment only. Some authors such as Karim and Zeger (1992), Breslow and Clayton (1993), and Lin and Breslow (1996), analyzed the data from each of the three experiments separately as well as the pooled data, where pooling was done ignoring the longitudinal dependence among the summer and the first fall data. Thus, to avoid any problems that may be caused by the longitudinal dependence, Sutradhar and Rao (2003) analyzed the data for 40 animals from the summer and the second fall experiments. For convenience, we reproduce here the data used by Sutradhar and Rao (2003). This reproduction is shown in Figure 5.1, where the symbol '$*$' indicates that the mating occurred and the '$\circ$' indicates that the mating did not occur. Note that in preparing Figure 5.1, we have reorganized the data from Table 14.3 of McCullagh and Nelder (1989) slightly so that the covariate values corresponding to a given response are easily recognized.

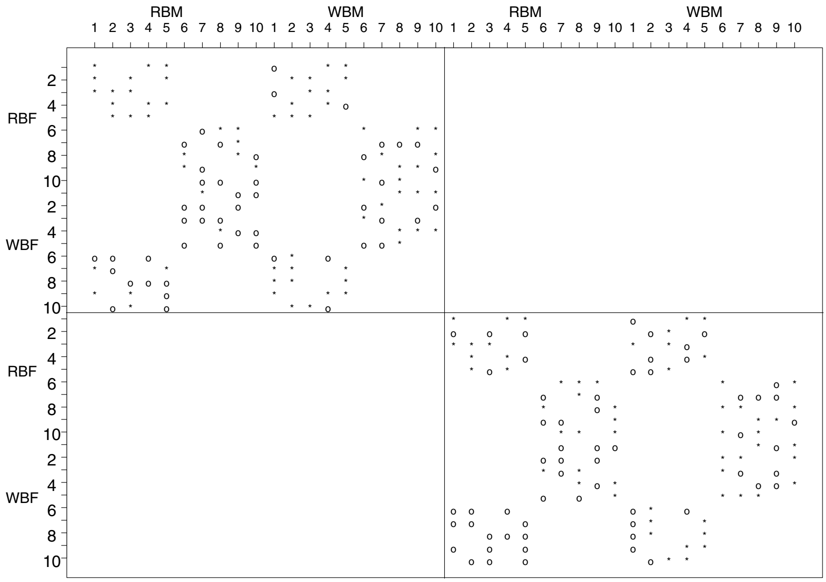


Fig. 5.1 Salamander mating data for summer and second fall experiments.

5.3.3.2 Binary Mixed Model for Salamander Data

For the data analysis, we have considered y_{ij} as the binary response for the mating of the ith female with the jth male $(i, j = 1, \ldots, 40)$, and $x'_{ij} = [x_{ij1}, x_{ij2}, x_{ij3}, x_{ij4}]$ be the corresponding 1×4 covariate vector, with

$$x_{ij1} = 1 \quad \text{for all i,j}$$

$$x_{ij2} = \begin{cases} 1 & \text{if the ith female belongs to WS group for any j} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ij3} = \begin{cases} 1 & \text{if the jth male belongs to the WS group for any i} \\ 0 & \text{otherwise} \end{cases}$$

and $x_{ij4} = x_{ij2}x_{ij3}$. The effects of these covariates are denoted by $\beta' = [\beta_1, \beta_2, \beta_3, \beta_4]$. Also note that as each animal was sequestered as pairs with six animals of the opposite sex, the six responses from these six animals will be structurally correlated as these responses are generated due to the common effect of the individual animal of the opposite sex. This common effect is considered to be a random effect in the present approach, and it is denoted by γ_i for the ith individual female, and by α_j

for the jth individual male. These random effects are assumed to have normal distributions with mean 0 and variances σ_γ^2 and σ_α^2, respectively. Consequently, in the notation of Section 5.3.1, one may write

$$E(Y_{ij}|\gamma_i^*, \alpha_j^*) = \pi_{ij}^*(\gamma_i^*, \alpha_j^*) = \frac{\exp(x_{ij}'\beta + \sigma_\gamma \gamma_i^* + \sigma_\alpha \alpha_j^*)}{1 + \exp(x_{ij}'\beta + \sigma_\gamma \gamma_i^* + \sigma_\alpha \alpha_j^*)}, \tag{5.151}$$

with $\gamma_i^* = \gamma_i/\sigma_\gamma$ and $\alpha_j^* = \alpha_j/\sigma_\alpha$.

5.3.3.3 Model Parameters Estimation and Interpretation

In connection with the estimation of the regression and the variance components of the present binary mixed model, some authors have used the method of moments. For example, we refer to the original analysis in McCullagh and Nelder (1989). The moment estimates in McCullagh and Nelder, in particular, the estimates of the variance components are, however, not consistent. See Kuk (1995, p. 404) for some discussions in this regard. As far as the regression estimates are concerned, they may or may not be consistent depending on the design matrix and sample size. But these estimates would be inefficient as the moment approach ignores the structural correlations among the responses in constructing the estimating equations for these parameters. Schall (1991), Breslow and Clayton (1993), Kuk (1995), and Lin and Breslow (1996) utilize these structural correlations indirectly, as they obtain the estimates of the regression and variance components, parameters by using the estimates of the random effects γ_i and α_j. More specifically, Schall (1991) and Breslow and Clayton (1993) [see also McGilchrist (1994)] consider certain adjustment to the asymptotically biased and inconsistent best linear unbiased prediction (BLUE) estimates for β and the random effects, and then estimate the variance components by using normal theory procedures. These estimators, however, still exhibit considerable bias particularly with regard to the variance components. Kuk (1995) and Lin and Breslow (1996) proposed independently certain bias correction procedures, but these procedures are known to be satisfactory for small values of the variance components or they produce large standard errors yielding large mean squared errors. As opposed to these procedures, Jiang (1998) introduced a simulated moment approach which always yields consistent estimators for the parameters of the mixed model. But as discussed in Section 5.2.5.4 in connection with a binary mixed model with a single component of dispersion, these moment estimates can be seriously inefficient. It was also shown that the GQL approach produces highly efficient estimates, the ML approach being optimal. Note, however, that in the present two-way factorial design setup, the ML inferences are extremely complicated. We, therefore, follow Sections 5.3.2.1 to 5.3.2.3, and obtain the GQL estimates for the parameters in (5.151). To be more specific, we now apply the three-step GQL estimation approach given in Section 5.3.2.4, to the salamander data set presented in Figure 5.1 for the estimation of the regression effects as well as the variance components of the female and male random effects. With initial values of

$\beta_1 = 0.02$, $\beta_2 = 0.10$, $\beta_3 = -0.03$, $\beta_4 = 0.04$, $\sigma_\gamma^2 = 0.5$, and $\sigma_\alpha^2 = 0.5$, a cyclical operation of the iterative equations (5.148), (5.149), and (5.150) yields the estimates along with their standard errors as shown in Table 5.4. In the same table, we also reproduce the so-called Gibbs estimates from Karim and Zeger (1992), and the PQL estimates from Lin and Breslow (1996).

Table 5.4 Estimates of regression (β) and variance components of the female (σ_γ^2) and male (σ_α^2) random effects and their estimated standard errors for the salamander data.

		Parameters					
Method		β_1	β_2	β_3	β_4	σ_γ^2	σ_α^2
GQL	Estimate	1.32	-3.25	-0.65	3.65	2.37	1.28
	ESE	1.13	0.99	0.48	0.93	0.72	0.51
PQL	Estimate	0.68	-2.16	-0.49	2.65	0.99	0.81
	ESE	0.37	0.55	0.43	0.64	-	-
Gibbs	Estimate	1.03	-3.01	-0.69	3.74	1.50	1.38
	ESE	0.43	0.60	0.50	0.68	-	-

Note that the standard errors of the variance component estimates for the PQL and Gibbs approaches were not available, whereas under the GQL approach, they were computed by using the standard formulas developed in (5.130) for the estimate of σ_γ^2, and in (5.141) for the estimate of σ_α^2. Further note that the GQL estimates of these variance components are quite different than those of Lin and Breslow (1996) and Karim and Zeger (1992).

With regard to the estimation of the regression parameters, the GQL estimates appear to be similar to Gibbs regression estimates of Karim and Zeger (1992), but standard errors are different. We must, however, caution the readers that the GQL estimates reported in Table 5.4 are in fact not directly comparable with the estimates of Karim and Zeger, and Lin and Breslow. This is because the latter authors have analyzed the pooled data (by pooling the summer, fall 1, and fall 2 data), whereas Sutradhar and Rao (2003) have analyzed the summer and fall 2 data only, in order to avoid longitudinal dependence of the summer and fall 1 data, as mentioned before.

Turning back to the GQL estimates, it is clear that the second and the fourth covariates appear to be highly significant. As $\hat{\beta}_2$ is negative, the mating occurrence rate for WSF appears to be small. This means that RBF has a larger mating occurrence rate as compared to the WSF. As the standard errors of the estimates of $\hat{\beta}_1$ (intercept parameter) and $\hat{\beta}_3$ appear to be relatively large as compared to their values, these covariates do not appear to be highly significant. A moderately negative large value of $\hat{\beta}_3 = -0.65$, nevertheless, indicates that RBM has a larger mating occurrence rate as compared to the WSM. Thus, in general, the salamanders from Rough Butt appear to have more mating occurrence rates as compared to the salamanders from White Side. But, the highly positive interaction (as compared to its standard error) indicates that the WSF and WSM have more mating occurrences among themselves.

Finally, a larger value of $\hat{\sigma}^2_\gamma$ as compared to the value of $\hat{\sigma}^2_\alpha$ indicates that irrespective of the locations, female salamanders appear to have high variability in matings as compared to the male salamanders.

5.4 Semiparametric Approach

Consider the binary mixed model $(5.1)-(5.3)$ and assume that the binary responses for the members of a given family are influenced by a random family effect, but unlike in Sections 5.1 – 5.3, we assume that the distribution of the random effects from independent families is unknown. Instead, the moments of the random effects $\gamma_i (i = 1, \ldots, K)$ up to order four are known and they are given by

$$E\gamma_i^r = \delta_r(\sigma^2_\gamma) = \sum_{s=1}^{r} c_{r,s}\sigma_\gamma^{r+1-s}, \text{ for } r = 1, \ldots, 4, \tag{5.152}$$

and

$$E\gamma_i^r = o(\sigma^r_\gamma), \text{ for } r \geq 5,$$

where $c_{r,s}$ are suitable known constants for $r = 1, \ldots, 4$.

5.4.1 GQL Estimation

By using the general exponential family density

$$f(y_{ij}|\eta_{ij}) = \exp[\{y_{ij}\eta_{ij} - a(\eta_{ij})\} + b(y_{ij})],$$

with $\eta_{ij} = \theta_{ij} + z_{i1}\gamma_i$, and $\theta_{ij} = x'_{ij}\beta$, it was shown in Section 4.4.2 how to construct the estimating equation for the regression effect β, when σ_γ is assumed to be known. To be specific, as it was given in Section 4.4.2, the estimating equation for β is given by

$$\sum_{i=1}^{K} \frac{\partial M'_{i,1}}{\partial \beta} M_{i,2}^{-1}(y_i - M_{i,1}) = 0, \tag{5.153}$$

where $M_{i,1}$ is the mean vector defined as

$$M_{i,1} = [M_{i1,1}, \ldots, M_{ij,1}, \ldots, M_{in_i,1}]',$$

with

$$M_{ij,1} = E[Y_{ij}] = a'_{ij} + \frac{\sigma^2_\gamma}{2} z^2_{i1} a'''_{ij} + \frac{\delta_3(\sigma^2_\gamma)}{6} z^3_{i1} a^{IV}_{ij} + \frac{\delta_4(\sigma^2_\gamma)}{24} z^4_{i1} a^{V}_{ij},$$

as in (4.131), and the covariance matrix $M_{i,2}$ can be computed by (4.134) and (4.135). Note that these mean vectors and covariance matrix may now easily be computed for the binary case by using the appropriate formula for a'_{ij} and further derivatives. In the binary case

$$a'_{ij}(\eta_{ij}) = \frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} = m_{ij} \text{ (say)}. \tag{5.154}$$

One may, consequently, derive the formulas for the higher-order derivatives up to order six as follows. Note that these formulas are needed to compute $M'_{i,1}$ and $M_{i,2}$ for the construction of the estimating equation (5.153). The formulas are:

$$\begin{aligned}
a''_{ij} &= m_{ij}(1-m_{ij}) \\
a'''_{ij} &= a''_{ij}[1-2m_{ij}] \\
a^{IV} &= a''_{ij}[1-6m_{ij}+6m_{ij}^2] \\
a^{V} &= a'''_{ij}[1-12m_{ij}+12m_{ij}^2] \\
a^{VI} &= a^{IV}_{ij}[1-12m_{ij}+12m_{ij}^2] - 12\{a'''_{ij}\}^2.
\end{aligned} \tag{5.155}$$

As far as the derivative $\partial M'_{i,1}/\partial\beta$ is concerned, it has the same formula as in the Poisson case (see Section 4.4.2). For convenience, we rewrite the formula as

$$\partial M'_{i,1}/\partial\beta = \begin{bmatrix} w_{i1}x'_{i1} \\ \vdots \\ w_{in_i}x'_{in_i} \end{bmatrix} = \begin{pmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & w_{in_i} \end{pmatrix} X_i$$

$$= W_i X_i, \text{ (say)},$$

where $X_i = [x_{i1}, \ldots, x_{ij}, \ldots, x_{in_i}]'$, and $W_i = \text{diag}[w_{i1}, \ldots, w_{ij}, \ldots, w_{in_i},]$ with

$$w_{ij} = \left[a''_{ij} + \frac{\sigma_\gamma^2}{2} z_{i1}^2 a^{IV}_{ij} + \frac{\delta^3(\sigma_\gamma^2)}{6} z_{i1}^3 a^{V}_{ij} + \frac{\delta_4(\sigma_\gamma^2)}{24} z_{i1}^4 a^{VI}_{ij}\right].$$

Let $\hat{\beta}_{GQL,SP}$ be the GQL estimator of β obtained by solving (5.153). It can be shown that asymptotically (as $K \to \infty$), for known σ_γ^2, the GQL estimator $\hat{\beta}_{GQL,SP}$ follows the multivariate Gaussian distribution with mean β and the covariance matrix given by

$$\text{cov}(\hat{\beta}_{GQL_{SP}}) = \text{limit}_{K\to\infty} \left[\sum_{i=1}^{K} \frac{\partial M'_{i,1}}{\partial\beta} M_{i,2}^{-1} \frac{\partial M_{i,1}}{\partial\beta'}\right]^{-1}. \tag{5.156}$$

5.4.2 A Marginal Quasi-Likelihood (MQL) Approach

Recall that the PQL approach suggested by Breslow and Clayton (1993) may yield inconsistent estimates, specially for σ_γ^2. Apart from PQL, these authors also have discussed a MQL approach which appears to be solving an estimating equation similar to (5.52) for the regression parameters. Under the normality assumption (same as for (5.52)) for the random effects, that is, $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma_\gamma^2)$, they have used an approximation to the mean vector π_i and the covariance matrix Σ_i. Thus, the estimate of β is bound to be worse or the same, in the sense of consistency and efficiency, as compared to the exact GQL estimate obtained from (5.52). In the next section, following Sutradhar and Rao (2001), we evaluate the performance of the MQL estimate of β with that of the semiparametric approach based estimate obtained from (5.153).

We now turn to the approximation of the mean vector and covariance matrix, used by Breslow and Clayton (1993) to construct the MQL estimating equation.

Approximation to the Mean Vector π and the Covariance Matrix Σ_i

Following Zeger et al. (1988), Breslow and Clayton (1993) approximated the mean vector π_i by

$$p_i^* = (p_{i1}^*, \ldots, p_{ij}^*, \ldots, p_{in_i}^*)',$$

where

$$p_{ij}^* = 1/[1 + \exp\{-c_{ij}' x_{ij}' \beta\}], \tag{5.157}$$

with $c_{ij} = (1 + c^2\sigma_\gamma^2)^{-1/2}$, and $c = 16(3^{1/2})/15\pi$. Similarly, the covariance matrix of y_i, that is, Σ_i, was approximated by

$$\Sigma_i^* = V_{i0} + \sigma_\gamma^2 V_{i0} U_{n_i} V_{i0}, \tag{5.158}$$

where U_{n_i} is the $n_i \times n_i$ unit matrix and

$$V_{i0} = \text{diag}[p_{i1}(\gamma_i = 0) q_{i1}(\gamma_i = 0), \ldots, p_{in_i}(\gamma_i = 0) q_{in_i}(\gamma_i = 0)],$$

with $p_{ij}(\gamma_i) = 1/[1 + \exp\{-x_{ij}'\beta - \gamma_i\}]$ and $q_{ij}(\gamma_i) = 1 - p_{ij}(\gamma_i)$.

MQL Estimating Equation for β

Now by using p_i^* for π_i, and Σ_i^* for Σ_i, into (5.52), one writes the MQL estimating equation for β as

$$\sum_{i=1}^{K} \frac{\partial p_i^{*'}}{\partial \beta} \Sigma_i^{*-1} (y_i - p_i^*) = 0. \tag{5.159}$$

Let $\hat{\beta}_{MQL}$ denote the estimate of β in this approach. Similar to (5.53), it then follows that $K^{1/2}(\hat{\beta}_{\text{MQL}} - \beta)$ is asymptotically multivariate normal with zero mean vector and covariance matrix V^*_β, which may be consistently estimated by

$$\hat{V}^*_\beta = \lim_{K\to\infty} K \left[\sum_{i=1}^{K} P^*_i \Sigma^{*}{}_i^{-1} P^{*'}{}_i \right]^{-1}_{\hat{\beta}_{\text{MQL}}}, \tag{5.160}$$

where $P^*_i = X'_i M^*_i C^*_i$, with

$$M^*_i = \text{diag}[p^*_{i1}q^*_{i1}, \ldots, p^*_{ij}q^*_{ij}, \ldots, p^*_{in_i}q^*_{in_i}] \text{ and } C^*_i = \text{diag}[c_{i1}, \ldots, c_{in_i}].$$

5.4.3 Asymptotic Efficiency Comparison: An Empirical Study

Note that the GQL estimating equation (5.153) is developed based on the assumption that the distribution of the random effect is unknown but its moments up to order four are known. Now to compare the efficiency of the normality based MQL estimates from (5.159) with that of the estimates from (5.153), we use the moments for normal distribution into (5.153). Thus, we put

$$\delta_1(\sigma^2_\gamma) = 0, \quad \delta_2(\sigma^2_\gamma) = 1, \quad \delta_3(\sigma^2_\gamma) = 0, \text{ and } \delta_4(\sigma^2_\gamma) = 3\sigma^4_\gamma$$

to construct the GQL estimating equation (5.153), and then calculate the asymptotic covariance of $\hat{\beta}_{GQL}$ by (5.156). Under the MQL approach we compute the asymptotic covariance of $\hat{\beta}_{MQL}$ by (5.160).

Now, to compute the relative efficiency of the MQL estimate for the uth $(u = 1, \ldots, p)$ regression component to the corresponding GQL estimate, we evaluate

$$\text{reff}(\hat{\beta}_{u(MQL)}) = v_{GQL}(u,u)/v_{MQL}(u,u), \tag{5.161}$$

where $v_{GQL}(u,u)$ and $v_{MQL}(u,u)$ are the uth diagonal elements of the covariance matrices, $\text{cov}(\hat{\beta}_{GQL})$ (5.156) and $\text{cov}(\hat{\beta}_{MQL})$ (5.160), respectively. In order to see how relative efficiency can vary with regard to the change in σ^2_γ values, we have computed the relative efficiency of $\hat{\beta}_{MQL}$ by (5.161) for two design matrices with $n_i = 6$ and $p = 2$ for $i = 1, \ldots, 100$. The two covariates under the first design (D_1) were chosen as

$$x_{ij1} = 1 \text{ for } j = 1, \ldots, 6;\ i = 1, \ldots, 100;$$

$$x_{ij2} = 1/j \text{ for } j = 1, \ldots, 6;\ i = 1, \ldots, 100;$$

and under the second design (D_2), they were:

$$x_{ij1} = 1 \text{ for } j = 1, \ldots, 6;\ i = 1, \ldots, 100;$$

$$x_{ij2} = \begin{cases} -1 & \text{for } j = 1, \ldots, 3;\ i = 1, \ldots, 50 \\ 0 & \text{for } j = 4, \ldots, 6;\ i = 1, \ldots, 50 \\ -1 & \text{for } j = 1, 2; \quad i = 51, \ldots, 100 \\ 0 & \text{for } j = 3, 4; \quad i = 51, \ldots, 100 \\ 1 & \text{for } j = 5, 6; \quad i = 51, \ldots, 100. \end{cases}$$

The relative efficiencies are reported in Table 5.5.

Table 5.5 Percentage relative efficiency of $\hat{\beta}_{MQL} = (\hat{\beta}_{1(MQL)}, \hat{\beta}_{2(MQL)})'$ to the GQL estimator $\hat{\beta}_{GQL} = (\hat{\beta}_{1(GQL)}, \hat{\beta}_{2(GQL)})'$ for selected values of σ_γ^2, and β_1, β_2.

			Values of σ_γ^2					
Design	Regression Coefficient	Relative Efficiency of	0.10	0.20	0.30	0.50	0.70	0.90
D_1	$\beta_1 = 1, \beta_2 = -1$	$\hat{\beta}_{1(MQL)}$	99	98	97	95	94	93
		$\hat{\beta}_{2(MQL)}$	99	98	97	95	93	91
	$\beta_1 = 0.25, \beta_2 = 0.25$	$\hat{\beta}_{1(MQL)}$	99	98	96	93	88	82
		$\hat{\beta}_{2(MQL)}$	99	98	97	94	92	89
	$\beta_1 = 0.25, \beta_2 = -0.25$	$\hat{\beta}_{1(MQL)}$	99	98	97	92	87	81
		$\hat{\beta}_{2(MQL)}$	99	98	97	94	91	87
D_2	$\beta_1 = 1, \beta_2 = -1$	$\hat{\beta}_{1(MQL)}$	99	99	98	97	96	95
		$\hat{\beta}_{2(MQL)}$	99	99	98	97	95	92
	$\beta_1 = 0.25, \beta_2 = 0.25$	$\hat{\beta}_{1(MQL)}$	99	97	95	89	81	73
		$\hat{\beta}_{2(MQL)}$	99	98	97	94	91	88
	$\beta_1 = 0.25, \beta_2 = -0.25$	$\hat{\beta}_{1(MQL)}$	99	97	94	88	81	72
		$\hat{\beta}_{2(MQL)}$	99	98	97	94	91	87

It is clear from the table that although the efficiency loss by the MQL approach is negligible for small values of $\sigma_\gamma^2 \leq 0.3$, the relative efficiency may, however, be quite low: as 72% for the intercept parameter and 87% for the slope parameter for $\sigma_\gamma^2 = 0.9$ under D_2. Under both designs, the relative efficiencies of the regression estimators appear to get smaller as σ_γ^2 gets larger, the situation being worse under D_2 as compared to D_1 for the intercept parameter. Under both designs, the efficiency loss appears to be significant even for moderate values of σ_γ^2 such as $\sigma_\gamma^2 = 0.5$, and 0.7. These relative efficiency results, therefore, indicate that the GQL approach leads to better regression estimates as compared to the MQL approaches, such as the MQL approach discussed in Breslow and Clayton (1993).

For the estimation of σ_γ^2 in the semiparametric approach, similar to the Poisson case, we refer to Sutradhar and Rao [2001, Section 4], among others.

5.5 Monte Carlo Based Likelihood Estimation

The computations for the Monte Carlo approach for the binary data are quite similar to those for the Poisson data discussed in Section 4.5. The only difference is that for the binary case one now uses the conditional density

$$f(y_{ij}|\gamma_i,\beta) = \exp[\{y_{ij}\eta_{ij} - a(\eta_{ij})\} + b(y_{ij})], \tag{5.162}$$

with $a(\eta_{ij}) = \log[1+\exp(\eta_{ij})]$, where $\eta_{ij} = \theta_{ij} + z_{i1}\gamma_i$, with $\theta_{ij} = x'_{ij}\beta$. This yields the likelihood for the data as

$$L(\beta,\sigma_\gamma^2) = \int \Pi_{i=1}^K f(y_i|\gamma_i) g_N(\gamma_i|\sigma_\gamma^2) d\gamma_i, \tag{5.163}$$

where $f(y_i|\gamma_i) = \Pi_{j=1}^{n_i} f(y_{ij}|\gamma_i)$, with $f(y_{ij}|\gamma_i)$ as in (5.162). One may then develop the Monte Carlo expectation-maximization (MCEM) approach for the binary data by replacing $f(\gamma_i|y_i)$ in Section 4.5.1 with binary density based $f(\gamma_i|y_i)$ from (5.163). In the same way, one may use the binary density based $f(\gamma_i|y_i)$ in Section 4.5.2 and develop the Monte Carlo Newton−Raphson (MCNR) approach for the binary data. With regard to the Monte Carlo expectation based Newton−Raphson iterative equation for the estimate of β, one now needs to use

$$\beta^{(r+1)} = \beta^{(r)} + E[\sum_{i=1}^K X'_i A_i(\gamma_i,\beta^{(r)})X_i]^{-1} E[\sum_{i=1}^K X'_i\{y_i - \pi_i^*(\gamma_i,\beta^{(r)})\}|y], \tag{5.164}$$

where

$$\pi_i^*(\gamma_i,\beta^{(r)}) = [\pi_{i1}^*(\gamma_i,\beta^{(r)}),\ldots,\pi_{ij}^*(\gamma_i,\beta^{(r)}),\ldots,\pi_{in_i}^*(\gamma_i,\beta^{(r)})]'$$

and

$$\begin{aligned} A_i(\gamma_i,\beta^{(r)}) = \text{diag}[&\pi_{i1}^*(\gamma_i,\beta^{(r)})\{1-\pi_{i1}^*(\gamma_i,\beta^{(r)})\},\ldots,\pi_{ij}^*(\gamma_i,\beta^{(r)})\{1-\pi_{ij}^*(\gamma_i,\beta^{(r)})\},\\ &\ldots,\pi_{in_i}^*(\gamma_i,\beta^{(r)})\{1-\pi_{in_i}^*(\gamma_i,\beta^{(r)})\}], \end{aligned} \tag{5.165}$$

with

$$\pi_{ij}^*(\gamma_i,\beta) = \frac{\exp(x'_{ij}\beta+\gamma_i)}{1+\exp(x'_{ij}\beta+\gamma_i)}.$$

Exercises

5.1. (Section 5.2.1) [Alternative expression for MM equations]
Write the moment equations in (5.28) and (5.29) as

$$\sum_{i=1}^{K}(w_i - \xi_i) = 0,$$

where $w_i = [w'_{i1}, w_{i2}]'$ and $\xi_i = [\xi'_{i1}, \xi_{i2}]'$, with

$$w_{i1} = \sum_{j=1}^{n_i} x_{ij}y_{ij},\ w_{i2} = \left(\sum_{j=1}^{n_i} y_{ij}\right)^2,$$

and

$$\xi_{i1} = \sum_{j=1}^{n_i} x_{ij}\pi_{ij}^{(s)}(\beta, \sigma_\gamma^2),\ \xi_{i2} = \left[\sum_{j=1}^{n_i} \pi_{ij}^{(s)}(\beta, \sigma_\gamma^2) + 2\sum_{j<k}^{n_i} \lambda_{ijk}^{(s)}(\beta, \sigma_\gamma^2)\right].$$

One then obtains the MM estimate of $\theta = [\beta', \sigma_\gamma^2]'$ by using the Gauss−Newton iterative equation

$$\hat{\theta}_{MM}(r+1) = \hat{\theta}_{MM}(r) + \left[\frac{\partial \sum_{i=1}^{K} \xi_i'}{\partial \theta}\right]_r^{-1} [\sum_{i=1}^{K}\{w_i - \xi_i\}]_r, \tag{5.166}$$

where $[]_r$ denotes that the expression within the square bracket is evaluated at $\theta = \hat{\theta}_{MM}(r)$, the estimate obtained for the rth iteration. It may then be shown that this moment estimator has the asymptotic variance given by

$$\mathrm{var}(\hat{\theta}_{MM}) = \mathrm{limit}_{K\to\infty} \left[\frac{\partial \sum_{i=1}^{K} \xi_i'}{\partial \theta}\right]^{-1} \sum_{i=1}^{K} V_i \left[\frac{\partial \sum_{i=1}^{K} \xi'}{\partial \theta}\right]^{-1}, \tag{5.167}$$

where $V_i = \mathrm{var}[W_i - \xi_i] = \mathrm{var}(W_i)$. Verify that this asymptotic variance is the same as the asymptotic variance given in (5.36).

5.2. (Section 5.2.5.3) [Aids to compute the elements of the information matrix in (5.92) under a special binary case]
For the binary logistic probability given by (5.81), and for $y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$ with $n_i = 4$, show that

$$J_i^{(s)} = N^{-1} \sum_{w=1}^{N} \exp(\gamma_{iw}^* \sigma_\gamma y_{i\cdot})(1 + \exp(x_i\beta + \sigma_\gamma \gamma_{iw}^*))^{-4}, \tag{5.168}$$

$$\begin{aligned} J_{i\beta}^{(s)} &= -4N^{-1}x_i \sum_{w=1}^{N} \exp\{x_i\beta + \gamma_{iw}^*\sigma_\gamma(1 + y_{i\cdot})\} \\ &\quad \times (1 + \exp(x_i\beta + \sigma_\gamma\gamma_{iw}^*))^{-5}, \end{aligned} \tag{5.169}$$

$$J_{i\beta\beta}^{(s)} = N^{-1}x_i^2 \sum_{w=1}^{N} \exp\{x_i\beta + \gamma_{iw}^*\sigma_\gamma(1 + y_{i\cdot})\}(1 + \exp(x_i\beta + \sigma_\gamma\gamma_{iw}^*))^{-6}$$

$$\times[4\{4\exp(x_i\beta+\sigma_\gamma\gamma^*_{iw})-1\}], \tag{5.170}$$

$$J^{(s)}_{i\sigma_\gamma} = N^{-1}\sum_{w=1}^{N}\gamma^*_{iw}\exp\{\gamma^*_{iw}\sigma_\gamma y_{i\cdot}\}(1+\exp(x_i\beta+\sigma_\gamma\gamma^*_{iw}))^{-5}$$

$$\times\left[y_{i\cdot}+\exp\{x_i\beta+\gamma^*_{iw}\sigma_\gamma\}(y_{i\cdot}-4)\right], \tag{5.171}$$

$$J^{(s)}_{i\beta\sigma_\gamma} = N^{-1}x_i\sum_{w=1}^{N}\gamma^*_{iw}\exp\{x_i\beta+\gamma^*_{iw}\sigma_\gamma(1+y_{i\cdot})\}(1+\exp(x_i\beta+\sigma_\gamma\gamma^*_{iw}))^{-6}$$

$$\times\left[4\{\exp(x_i\beta+\sigma_\gamma\gamma^*_{iw})(4-y_{i\cdot})-(y_{i\cdot}+1)\}\right], \tag{5.172}$$

and

$$J^{(s)}_{i\sigma_\gamma\sigma_\gamma} = N^{-1}\sum_{w=1}^{N}{\gamma^*}^2_{iw}\exp\{x_i\beta+\gamma^*_{iw}\sigma_\gamma(1+y_{i\cdot})\}(1+\exp(x_i\beta+\sigma_\gamma\gamma^*_{iw}))^{-6}$$

$$\times\left[\{y_{i\cdot}+\exp(x_i\beta+\sigma_\gamma\gamma^*_{iw})(4-y_{i\cdot})\}\{\exp(-x_i\beta-\sigma_\gamma\gamma^*_{iw})-4\}\right.$$

$$\left.+(y_{i\cdot}-4)(1+\exp(-x_i\beta-\sigma_\gamma\gamma^*_{iw})\right]. \tag{5.173}$$

5.3. (Section 5.3.2.2) [Derivative matrix for the GQL estimating equation (5.123)] Verify that the $((mn(n+1))/2)\times 1$ derivative matrix $B^{*(s)}$ in (5.123) can be computed by exploiting the formulas for the derivatives of two general elements

$$\pi^{(s)}_{ij}(\beta,\sigma^2_\gamma,\sigma^2_\alpha) \text{ and } \lambda^{(s)}_{iijk}(\beta,\sigma^2_\gamma,\sigma^2_\alpha)$$

in (5.101) and (5.104), respectively, with respect to σ^2_γ. Also verify that these derivatives are given by

$$\frac{\partial\pi^{(s)}_{ij}(\beta,\sigma^2_\gamma,\sigma^2_\alpha)}{\partial\sigma^2_\gamma} = \frac{1}{2\sigma_\gamma}\frac{1}{N}\sum_{w=1}^{N}\pi^*_{ij}(\gamma^*_{iw},\alpha^*_{jw})[1-\pi^*_{ij}(\gamma^*_{iw},\alpha^*_{jw})],$$

for $i=1,\ldots,m,\ j=1,\ldots,n$, and

$$\frac{\partial\lambda^{(s)}_{(ii)jk}(\beta,\sigma^2_\gamma,\sigma^2_\alpha)}{\partial\sigma^2_\gamma} = \frac{1}{2\sigma_\gamma}\frac{1}{N}\sum_{w=1}^{N}\pi^*_{ij}(\gamma^*_{iw},\alpha^*_{jw})\pi^*_{ik}(\gamma^*_{iw},\alpha^*_{kw})$$
$$\times[2-\pi^*_{ij}(\gamma^*_{iw},\alpha^*_{jw})-\pi^*_{ik}(\gamma^*_{iw},\alpha^*_{kw})],$$

for $i=1,\ldots,m,\ j<k,\ j,k=1,\ldots,n$.

5.4. (Section 5.3.2.3) [Derivative matrix for the GQL estimating equation (5.141)] Verify that the $((nm(m+1)/2)\times 1$ derivative matrix $\tilde{B}^{(s)}$ in (5.141) can be computed by exploiting the formulas for the derivatives of two general elements $\pi_{ij}^{(s)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)$ in (5.101) and $\lambda_{ir(jj)}^{(s)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)$ in (5.106), with respect to σ_α^2. Also verify that these derivatives are given by

$$\frac{\partial \pi_{ij}^{(s)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)}{\partial \sigma_\alpha^2} = \frac{1}{2\sigma_\alpha} \frac{1}{N} \sum_{w=1}^{N} \pi_{ij}^*(\gamma_{iw}^*, \alpha_{jw}^*)[1 - \pi_{ij}^*(\gamma_{iw}^*, \alpha_{jw}^*)],$$

for $i = 1, \ldots, m, \; j = 1, \ldots, n$, and

$$\frac{\partial \lambda_{ir(jj)}^{(s)}(\beta, \sigma_\gamma^2, \sigma_\alpha^2)}{\partial \sigma_\alpha^2} = \frac{1}{2\sigma_\alpha} \frac{1}{N} \sum_{w=1}^{N} \pi_{ij}^*(\gamma_{iw}^*, \alpha_{jw}^*) \pi_{rj}^*(\gamma_{rw}^*, \alpha_{jw}^*) \times [2 - \pi_{ij}^*(\gamma_{iw}^*, \alpha_{jw}^*) - \pi_{rj}^*(\gamma_{rw}^*, \alpha_{jw}^*)],$$

for $j = 1, \ldots, n, \; i < r, \; i, r = 1, \ldots, m$.

References

1. Breslow, N, E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Stat. Assoc.,* 88, 9 – 25.
2. Cohen, B. H. (1980). Chronic obstructive pulmonary disease: A challenge in genetic epidemiology. *Am. J. Epidemiol.,* 112, 274 – 288.
3. Ekholm, A., Smith, P. W. F., and McDonald, J. W. (1995). Marginal regression analysis of a multivariate binary response. *Biometrika*, 82, 847 – 854.
4. Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.,* 93, 720 – 729.
5. Jiang, J. & Zhang, W. (2001). Robust estimation in generalized linear mixed models. *Biometrika,* 88, 753 – 765.
6. Karim , M. R. & Zeger, S. L. (1992). Generalized linear models with random effects: Salamander mating revisited. *Biometrics,* 48, 631 – 644.
7. Kuk, A. Y. C. (1995). Asymptotically unbiased estimation in generalized linear models with random effects. *J. R. Statist. Soc.* B, 58, 619 – 678.
8. Lee, Y. & Nelder, J. (1996). Hierarchical generalized linear models. *J. R. Statist. Soc. B,* 58, 619 – 678.
9. Lin, X. & Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Am. Statist. Assoc.,* 91, 1007 – 1016.
10. Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992), Multivariate Regression Analysis for Categorical Data. *J. Roy. Statist. Soc. Ser. B*, 54, 3 – 40.
11. Mardia, K. V., Kent. J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press Inc..
12. McCullagh, P., & Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.). London: Chapman and Hall.
13. McGilchrist, C. A. (1994). Estimation in generalised linear mixed models. *J. R. Statist. Soc.* B, 56, 61 – 69.
14. Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika,* 78, 719 – 27.

15. Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statist. Sci.,* 18, 377 – 393.
16. Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhya,* 66, 261 – 289.
17. Sutradhar, B. C. & Mukerjee, R. (2005). On likelihood inference in binary mixed model with an application to COPD data. *Computat. Statist. Data Anal.,* 48, 345 – 361.
18. Sutradhar, B. C. & Rao, R. Prabhakar (2001). On marginal quasi-likelihood inference in generalized linear mixed models. *J. Multivar. Anal.,* 76, 1 – 34.
19. Sutradhar, B. C. & Rao, R. Prabhakar (2003). On quasilikelihood inference in generalized linear mixed models with two variance components. *Canad. J. Statist.,* 31, 415 – 435.
20. Ten Have, T. R. & Morabia, A. (1999). Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometrics,* 55, 85 – 93.
21. Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, 44, 1049 – 1060.

Appendix

Tables 5A−5E: COPD data. [Code: column 1 (C1)-Sibling identification; C2-IPF status (1 for without IPF, 0 for with IPF); C3-Intercept; C4-Gender (1 for male, 0 for female); C5-Race (1 for black, 0 for white); C6-Age (centered at 50) ; C7-Smoking status (1 for smoking, 0 for nonsmoking)].

Table 5A. COPD data from the siblings of 48 patients each with one sibling.

C1	C2	C3	C4	C5	C6	C7
10005	0	1	1	0	-6	1
10007	0	1	1	1	9	1
10023	0	1	1	1	7	0
10024	0	1	1	1	-3	0
10031	1	1	1	0	11	0
10032	0	1	0	1	-23	0
10033	0	1	0	0	4	1
10040	0	1	1	0	-14	1
10041	1	1	0	0	-2	0
10050	0	1	0	1	-10	0
10053	0	1	1	0	-19	0
10063	0	1	1	0	3	1
10068	0	1	1	0	-4	1
10069	0	1	0	1	-12	0
10081	0	1	0	0	10	1
10088	0	1	1	0	-34	1
10091	1	1	1	1	15	0
10102	0	1	1	0	-12	0
10105	0	1	0	0	4	0
10113	1	1	1	1	13	0
10117	0	1	0	1	-37	0
10124	0	1	0	0	4	1
10130	1	1	0	0	1	1
10134	0	1	1	0	2	1
10137	1	1	0	0	0	0
10141	1	1	1	0	6	0
10155	0	1	1	1	-7	1
10160	0	1	1	1	23	0
10162	1	1	1	0	2	1
10173	0	1	0	1	4	0
10189	0	1	1	0	0	0
10196	0	1	1	1	-22	0
10198	0	1	1	0	-33	0
10202	0	1	1	0	13	1
10204	0	1	1	1	-9	0
10212	0	1	0	0	11	0
10213	1	1	0	1	14	1
10198	0	1	1	0	-33	0
10202	0	1	1	0	13	1
10204	0	1	1	1	-9	0
10212	0	1	0	0	11	0
10213	1	1	0	1	14	1
10220	0	1	1	1	6	1
10230	0	1	0	1	0	0
10235	1	1	1	1	0	1
10237	0	1	1	0	-10	1
10249	0	1	1	0	-6	1
10252	1	1	1	1	8	1
10255	1	1	1	0	-5	1
10260	0	1	0	0	-10	0
10263	0	1	0	0	-5	0
10264	1	1	0	0	-5	1
15001	0	1	1	1	-28	1

Table 5B. COPD data from the siblings of 23 patients each with two siblings.

C1	C2	C3	C4	C5	C6	C7
10009	1	1	0	0	11	1
10009	0	1	1	0	14	0
10010	0	1	1	0	10	1
10010	0	1	1	0	9	1
10011	1	1	0	0	9	0
10011	0	1	1	0	7	1
10058	0	1	0	1	1	1
10058	0	1	1	1	5	0
10074	0	1	0	0	-2	0
10074	0	1	1	0	-6	0
10078	0	1	0	1	-19	0
10078	0	1	1	1	-23	1
10084	0	1	1	0	-16	1
10084	0	1	1	0	-22	1
10089	1	1	1	0	8	1
10089	1	1	1	0	2	1
10092	0	1	1	0	-36	0
10092	0	1	1	0	-39	0
10138	0	1	0	0	1	0
10138	0	1	1	0	4	1
10151	0	1	0	0	-8	0
10151	0	1	1	0	-17	0
10154	0	1	0	0	-4	0
10154	0	1	1	0	1	0
10166	1	1	0	0	-8	0
10166	0	1	1	0	7	0
10178	0	1	0	0	5	0
10178	0	1	1	0	10	1
10186	0	1	1	0	-10	1
10186	0	1	1	0	-13	1
10200	1	1	0	1	5	0
10200	0	1	1	1	16	1
10203	0	1	0	1	1	0
10203	0	1	0	1	-4	0
10226	0	1	0	1	-11	1
10226	0	1	1	1	-1	1
10231	0	1	0	1	-34	0
10231	0	1	1	1	-33	0
10241	1	1	0	0	-8	1
10241	0	1	1	0	1	0
10250	0	1	1	1	4	1
10250	0	1	1	1	-4	0
10254	1	1	0	1	7	1
10254	0	1	1	1	19	0
10257	1	1	1	1	18	1
10257	0	1	1	1	-1	0

Table 5C. COPD data from the siblings of 17 patients each with 3 siblings.

C1	C2	C3	C4	C5	C6	C7
10016	1	1	0	0	22	0
10016	0	1	0	0	5	0
10016	0	1	1	0	12	0
10018	0	1	0	1	3	0
10018	1	1	1	1	11	1
10018	0	1	1	1	6	0
10061	0	1	0	0	9	0
10061	1	1	0	0	7	0
10061	1	1	1	0	5	0
10083	1	1	0	1	5	1
10083	0	1	0	1	2	1
10083	0	1	1	1	11	1
10098	1	1	0	0	4	0
10098	1	1	1	0	12	0
10098	1	1	1	0	5	1
10128	1	1	0	1	-2	0
10128	0	1	0	1	-3	1
10128	1	1	1	1	3	0
10136	0	1	0	0	-1	1
10136	0	1	1	0	-7	0
10136	0	1	1	0	-16	1
10140	0	1	0	0	10	1
10140	0	1	0	0	1	1
10140	0	1	1	0	6	0
10153	0	1	0	1	1	1
10153	0	1	1	1	2	0
10153	0	1	1	1	-4	0
10159	1	1	0	0	5	0
10159	0	1	0	0	2	0
10159	0	1	1	0	7	1
10168	1	1	0	0	-9	1
10168	1	1	1	0	-11	1
10168	1	1	1	0	-13	1
10185	0	1	0	0	15	0
10185	0	1	1	0	6	0
10185	1	1	1	0	3	0
10188	0	1	0	0	-37	0
10188	0	1	0	0	-38	0
10188	0	1	1	0	-34	0
10192	0	1	0	1	12	1
10192	0	1	0	1	7	0
10192	0	1	0	1	5	1
10207	1	1	0	0	10	0
10207	0	1	0	0	-5	0
10207	1	1	1	0	8	0
10210	0	1	0	1	-7	0
10210	0	1	1	1	18	1
10210	1	1	1	1	5	1
10229	0	1	1	0	2	0
10229	1	1	1	0	-1	0
10229	0	1	1	0	-4	0

Table 5D. COPD data from the siblings of 7 patients each with four siblings.

C1	C2	C3	C4	C5	C6	C7
10028	0	1	0	1	8	0
10028	0	1	0	1	1	0
10028	0	1	0	1	-3	1
10028	0	1	1	1	-12	1
10039	0	1	0	0	-2	1
10039	0	1	0	0	-4	0
10039	1	1	0	0	-6	1
10039	0	1	0	0	-9	0
10051	1	1	0	0	2	1
10051	0	1	1	0	10	0
10051	0	1	1	0	-2	0
10051	1	1	1	0	-4	1
10090	0	1	0	1	-18	0
10090	0	1	0	1	-19	0
10090	0	1	0	1	-25	0
10090	0	1	1	1	-27	0
10209	0	1	0	1	-11	1
10209	0	1	1	1	-5	0
10209	0	1	1	1	-8	0
10209	0	1	1	1	-19	1
10242	0	1	0	1	0	1
10242	0	1	0	1	-7	1
10242	0	1	0	1	-7	1
10242	1	1	1	1	-4	1
10251	0	1	0	0	9	0
10251	1	1	0	0	4	0
10251	0	1	1	0	6	0
10251	0	1	1	0	-1	0

Table 5E. COPD data from the siblings of 5 patients each with six siblings.

C1	C2	C3	C4	C5	C6	C7
10003	1	1	0	1	8	1
10003	1	1	0	1	5	1
10003	0	1	0	1	-9	1
10003	1	1	1	1	9	1
10003	0	1	1	1	2	1
10003	1	1	1	1	-6	1
10022	0	1	0	0	-8	0
10022	0	1	0	0	-10	0
10022	0	1	1	0	10	1
10022	0	1	1	0	6	1
10022	0	1	1	0	2	0
10022	0	1	1	0	-3	1
10095	1	1	0	1	-13	1
10095	1	1	0	1	-21	1
10095	0	1	1	1	-8	1
10095	0	1	1	1	-11	1
10095	0	1	1	1	-22	1
10095	0	1	1	1	-24	0
10158	1	1	0	0	-12	0
10158	1	1	0	0	-15	1
10158	1	1	1	0	4	1
10158	1	1	1	0	-5	0
10158	1	1	1	0	-6	1
10158	1	1	1	0	-9	1
10169	1	1	0	0	0	1
10169	1	1	0	0	-4	1
10169	1	1	0	0	-6	0
10169	0	1	1	0	7	0
10169	0	1	1	0	2	1
10169	0	1	1	0	-1	1

Chapter 6
Longitudinal Models for Count Data

In longitudinal studies for count data, a small number of repeated count responses along with a set of multidimensional covariates are collected from a large number of independent individuals. For example, in a health care utilization study, the number of visits to a physician by a large number of independent individuals may be recorded annually over a period of several years. Also, the information on the covariates such as gender, number of chronic conditions, education level, and age, may be recorded for each individual. For $i = 1, \ldots, K$, and $t = 1, \ldots, T$, let y_{it} denote the count response and $x_{it} = (x_{it1}, \ldots, x_{itp})'$ denote the p-dimensional covariate vector collected at time point t from the ith individual. Let β be the effect of x_{it} on y_{it}. Note that because $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are T repeated count responses from the same individual, it is most likely that they are autocorrelated. The scientific concern is to find β, the effects of the covariates on the repeated count responses, after taking their autocorrelations into account.

Note that there are situations in practice, where the covariates of the ith individual may be time independent. We denote such covariates by $\tilde{x}_i = (x_{i1}, \ldots, x_{ip})'$. This is a simpler special case of the general situation with time-dependent covariates x_{it}. In Section 6.1, we provide the marginal distributional properties of the count response variable Y_{it} under the general situation when corresponding covariates are time dependent. For simplicity, Section 6.2 discusses the estimation of β by pretending that the repeated count responses are independent, even though in reality they are autocorrelated. In Section 6.3, we provide several autocorrelation structures for the repeated count data for the special case with time-independent covariates. A unified generalized quasi-likelihood (GQL) approach is discussed in Section 6.4 for the estimation of the regression effects β after taking the stationary correlations of the data into account.

Note that stationary autocorrelation models can be generalized to the nonstationary cases in various ways. We consider two types of nonstationary models. First, we consider a class of nonstationary autocorrelation models where all models produce the same specified marginal mean and variance functions. These models are given in Section 6.5. The same section also contains the estimating equation for β after taking the nonstationary correlations into account. Second, in Section 6.6, we

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_6,

demonstrate that the stationary autocorrelation models discussed in Section 6.3 may be generalized to a nonstationary class of models where these models may produce different marginal means and variances along with different correlation structures. The inferences for the regression effects β, after taking the nonstationary correlation structure of the repeated data into account are discussed in details, including the model misspecification effects. Note that in the stationary case, model selection is not necessary for the estimation of the regression effects, whereas model selection becomes an important issue in the nonstationary case. This model selection problem is also discussed in Section 6.6 for the second type of nonstationary autocorrelation models. A data example is considered in Section 6.7 to illustrate both correlation model selection and estimation of the parameters.

6.1 Marginal Model

Suppose that each of the count response variables $Y_{i1},\ldots,Y_{it},\ldots,Y_{iT}$ for the ith $(i=1,\ldots,K)$ follows the well-known Poisson distribution with a suitable mean parameter. Let $\mu_{it}=\exp(x'_{it}\beta)$ denote the mean of the Poisson distribution for Y_{it}. In the form of exponential density, one may then write the marginal distribution of Y_{it} as

$$f(y_{it})=\exp[\{y_{it}\theta_{it}-a(\theta_{it})\}+b(y_{it})] \tag{6.1}$$

[Nelder and Wedderburn (1972)], with

$$\theta_{it}=x'_{it}\beta,;a(\theta_{it})=\exp(\theta_{it}),\ \text{and}\ b(y_{it})=\log(\frac{1}{y_{it}!}).$$

We denote this marginal Poisson distribution as $Y_{it}\sim \text{Poi}(\mu_{it})$. For an auxiliary parameter s, by using the moment generating function (m.g.f.) of Y_{it} [see (4.9), also Exercise (4.5)] given by

$$M_{Y_{it}}(s)=E[\exp(sY_{it})]=\exp[a(s+\theta_{it})-a(\theta_{it})], \tag{6.2}$$

one may obtain the basic properties such as the first four moments of the marginal distribution (6.1) as in the following lemma.

Lemma 6.1 The first four moments of Y_{it} under the exponential family density (6.1) are given by

$$\mu_{it}=[Y_{it}]=a'(\theta_{it})$$

$$\sigma_{itt}=\text{var}[Y_{it}]=a''(\theta_{it})$$

$$\tilde{\delta}_{itt}=E[Y_{it}-\mu_{it}]^3=a'''(\theta_{it})$$

$$\tilde{\phi}_{itttt}=E[Y_{it}-\mu_{it}]^4=a''''(\theta_{it})+3\sigma^2_{itt}, \tag{6.3}$$

where $a'(\theta_{it})$, $a''(\theta_{it})$, $a'''(\theta_{it})$, and $a''''(\theta_{it})$ are, respectively, the first−, second−, third−, and the fourth-order derivatives of $a(\theta_{it})$ with respect to θ_{it}.

In the present longitudinal setup, the repeated count responses $y_{i1},\ldots,y_{it},\ldots,y_{iT}$ are most likely to be correlated, and these correlations, unlike the familial correlations developed through random effects in Chapter 4, should reflect the time effects. Some suitable modelling for this type of time effects based correlations is discussed in Section 6.3 for the cases when covariates are stationary (i.e., time independent), and in Sections 6.5 and 6.6 when covariates are nonstationary (i.e., time dependent). Note that if one is, however, interested to obtain only a consistent estimate for β as opposed to a consistent as well as efficient estimate, then, the repeated responses may be treated as independent and the marginal distribution (6.1) or the marginal properties in Lemma 6.1 may be exploited to construct suitable estimating equations to achieve such a goal. In the following section, we discuss three standard marginal model based estimation techniques that use either the marginal density in (6.1) or only the first two moments from Lemma 6.1.

6.2 Marginal Model Based Estimation of Regression Effects

Method of Moments (MM): Irrespective of the cases whether the repeated counts $y_{i1},\ldots,y_{it},\ldots,y_{iT}$ are independent or autocorrelated, one may always obtain the moment estimate of β by solving the moment equation

$$\sum_{i=1}^{K}\sum_{t=1}^{T}[x_{it}(y_{it}-a'(\theta_{it}))]=0, \tag{6.4}$$

where $a'(\theta_{it})=\mu_{it}=\exp(x'_{it}\beta)$ for Poisson y_{it}. By writing $y_i=(y_{i1},\ldots,y_{it},\ldots,y_{iT})'$: $T\times 1$; $\mu_i=(\mu_{i1},\ldots,\mu_{it},\ldots,\mu_{iT})' : T\times 1$; and $X_i=(x_{i1},\ldots,x_{it},\ldots,x_{iT})' : T\times p$, the moment equation (6.4) may be re-expressed as

$$\sum_{i=1}^{K}[X'_i(y_i-\mu_i)]=0. \tag{6.5}$$

Let the moment estimator of β, the root of the moment equation (6.5), be denoted by $\hat{\beta}_M$. This root may be obtained by using the iterative equation

$$\hat{\beta}_M(r+1)=\hat{\beta}_M(r)+\left[\sum_{i=1}^{K}X'_iA_iX_i\right]^{-1}_{(r)}\left[\sum_{i=1}^{K}X'_i(y_i-\mu_i)\right]_{(r)}, \tag{6.6}$$

where $A_i=\text{diag}[a''(\theta_{it})]=\text{diag}[\sigma_{itt}]$, and $[\cdot]_{(r)}$ denotes that the expression within the brackets is evaluated at $\beta=\hat{\beta}_M(r)$, the rth iterative value for $\hat{\beta}_M$. Note that because (6.5) is an unbiased estimating equation for the zero vector, $\hat{\beta}_M$ is a consistent estimator. Furthermore, because K individuals are chosen independently, by using multivariate central limit theorem [Mardia, Kent and Bibby (1979, p. 51)] it follows from (6.6) that $K^{\frac{1}{2}}(\hat{\beta}_M - \beta)$ is asymptotically multivariate Gaussian with zero mean vector and covariance matrix V_M given by

$$V_M = \text{limit}_{K\to\infty} K \left[\sum_{i=1}^{K} X_i' A_i X\right]^{-1} \left[\sum_{i=1}^{K} X_i' A_i^{1/2} C_i A_i^{1/2} X_i\right] \left[\sum_{i=1}^{K} X_i' A_i X\right]^{-1}, \quad (6.7)$$

where C_i is the true correlation matrix of y_i, which may be unknown. This covariance matrix V_M may, however, be estimated by using the sandwich type estimator

$$\hat{V}_M = \text{limit}_{K\to\infty} K \left[\sum_{i=1}^{K} X_i' A_i X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i'(y_i - \mu_i)(y_i - \mu_i)' X_i\right] \left[\sum_{i=1}^{K} X_i' A_i X_i\right]^{-1} \quad (6.8)$$

[see for example, Liang and Zeger (1986, p. 15)].

Quasilikelihood (QL) Method : Note that when there is a functional relationship between the mean and the variance of the response, Wedderburn (1974) [see also McCullagh (1983)] proposed a QL approach for independent data which exploits both mean and the variance in estimating the regression effects β. The QL estimating equation for β is given by

$$\sum_{i=1}^{K}\sum_{t=1}^{T}\left[\frac{\partial a'(\theta_{it})}{\partial \beta}\frac{(y_{it} - a'(\theta_{it}))}{\text{var}(y_{it})}\right] = 0, \quad (6.9)$$

where the $\text{var}(Y_{it}) = a''(\theta_{it})$ is a function of the mean parameter $a'(\theta_{it}) = \mu_{it}$. In the Poisson case, for example,

$$\text{var}(Y_{it}) = a''(\theta_{it}) = a'(\theta_{it}) = \mu_{it} = \exp(x_{it}'\beta).$$

Notice that there is no difference between this QL estimating equation (6.9) and the MM estimating equation (6.4).

We remark, however, that as opposed to the independence case, in a practical situation one would also exploit the correlation properties of the repeated responses in generalizing the QL estimating equation (6.9), but the MM approach will still use the estimating equation (6.5). Thus, in the longitudinal setup, the generalized QL approach will yield a different estimate for β than the MM approach.

Marginal Likelihood (ML) Method: It is true that the repeated counts

$$y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$$

are autocorrelated. If the correlations are, however, ignored, that is, the repeated responses are treated to be independent, then one may maximize the marginal likelihood function to obtain an independence assumption based 'working' likelihood estimate of β. By (6.1), the log of the marginal likelihood function of β is given by

$$\log L(\beta) = \sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})], \tag{6.10}$$

yielding the likelihood equation for β as

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it} - a'(\theta_{it})]\frac{\partial \theta_{it}}{\partial \beta} = 0. \tag{6.11}$$

Because, $\theta_{it} = x'_{it}\beta$, this likelihood equation is the same as the MM equation (6.4). Thus it is clear that all three approaches, namely, the MM, QL and ML methods yield the same estimate for β. All three approaches yield a consistent estimate for this regression effect.

6.3 Correlation Models for Stationary Count Data

Note that a marginal model based estimation approach may not yield an efficient regression estimate. Obtaining an efficient estimate will require exploitation of the joint probability or correlation model for the repeated count data. In this section, we discuss this issue, for a simpler situation when covariates of an individual are time independent. Note that this situation can arise in some longitudinal studies such as in a longitudinal clinical study where, for example, the number of weekly asthma attacks is recorded as the responses over a small period such as four weeks of time. Here, it is likely that the covariate information such as gender, education level, and number of other chronic diseases of the individual will remain the same for each week for the duration of the study over four weeks. This is, however, true that the repeated responses will still be correlated due to the influence of time, the time being a stochastic factor. In the end, it is of main interest to find the effects of the covariates on the responses after taking the correlations of the responses into account.

Recall that $\tilde{x}_i = (x_{i1}, \ldots, x_{ip})'$ denote the time-independent covariate vector for the ith individual. For this time-independent covariate, the mean and the variance of y_{it} may be written, following Lemma 6.1, as

$$E[Y_{it}] = \text{var}[Y_{it}] = \tilde{\mu}_i = \exp(\tilde{x}'_i\beta), \tag{6.12}$$

yielding the mean vector and the diagonal matrix of the variances as

$$\mu_i = \tilde{\mu}_i\mathbf{1},\ A_i = \text{diag}(\sigma_{itt}) = \text{diag}(\tilde{\mu}_i), \tag{6.13}$$

where $\mathbf{1}$ is the $T \times 1$ unit vector.

As far as the correlation structures for the repeated counts $y_{i1}, \ldots, y_{iT}$ are concerned, it was speculated in some of the original studies such as in Liang and Zeger (1986) that the correlations of the repeated data may follow Gaussian type such as autoregressive order 1 (AR(1)), moving average order (1) (MA(1)), or exchangeable (equicorrelations) correlation structures. But, as it is not easy to know the underlying true correlation structure, these authors have used a 'working' correlation structure based generalized estimating equations (GEE) approach for the efficient estimation of the regression effects. We discuss this GEE approach and its serious limitations in Section 6.4.

We now provide three correlation models [Sutradhar (2003), McKenzie (1988)] that yield the speculated AR(1), MA(1), and equicorrelation structures for repeated count data. In fact, these three low-order models are easily extendable to other possible higher-order models such as AR(2), MA(2), and ARMA(1,1) models.

6.3.1 Poisson AR(1) Model

Let $y_{i1} \sim \text{Poi}(\tilde{\mu}_i)$, where $\tilde{\mu}_i = \exp(\tilde{x}_i'\beta)$ as in (6.12). Furthermore, for $t = 2,\ldots,T$, let the response y_{it} at time t be related to $y_{i,t-1}$ at time $t-1$ as

$$y_{it} = \rho * y_{i,t-1} + d_{it}, \tag{6.14}$$

[McKenzie (1988), Sutradhar (2003)] where it is assumed that for given $y_{i,t-1}$, $\rho * y_{i,t-1}$ denotes the so-called binomial thinning operation (McKenzie, 1988). That is,

$$\begin{aligned} \rho * y_{i,t-1} &= \sum_{j=1}^{y_{i,t-1}} b_j(\rho) \\ &= z_{i,t-1}, \text{say}, \end{aligned} \tag{6.15}$$

with $\Pr[b_j(\rho) = 1] = \rho$ and $\Pr[b_j(\rho) = 0] = 1 - \rho$. Furthermore, it is assumed in (6.14) that $d_{it} \sim P(\tilde{\mu}_i(1-\rho))$ and is independent of $z_{i,t-1}$.

It then follows that each y_{it} satisfying the model (6.14) has marginally Poisson distribution with parameters as in (6.12). Also by direct calculation, it can be shown that

$$\begin{aligned} E[Y_{it}] &= E_{Y_{i,t-1}} E[Y_{it}|Y_{i,t-1}] = \tilde{\mu}_i \\ \text{var}[Y_{it}] &= E_{Y_{i,t-1}} \text{var}[Y_{it}|Y_{i,t-1}] + \text{var}_{Y_{i,t-1}} E[Y_{it}|Y_{i,t-1}] = \tilde{\mu}_i. \end{aligned} \tag{6.16}$$

Next, by similar calculations as in (6.16), for lag $\ell = 1,\ldots,T-1$, it can be shown from (6.14) that $E(Y_{it}Y_{i,t-\ell}) = \tilde{\mu}_i^2 + \tilde{\mu}_i\rho^\ell$, yielding the lag ℓ correlation between y_{it} and $y_{i,t-\ell}$, say $c^*_{i,(t-\ell)t}(\rho)$, as

$$\begin{aligned} \text{corr}(Y_{it}, Y_{i,t-\ell}) &= c^*_{i,(t-\ell)t}(\rho) \\ &= \rho^\ell, \end{aligned} \tag{6.17}$$

which is the same as lag ℓ correlation under the Gaussian AR(1) autocorrelation structure. But, the ρ parameter under the present AR(1) model (6.14) must satisfy the range restriction $0 \leq \rho \leq 1$, whereas in the Gaussian AR(1) structure ρ lies in the range $-1 < \rho < 1$.

6.3.2 *Poisson MA(1) Model*

For a scale parameter ρ, let

$$d_{it} \overset{\text{iid}}{\sim} \text{Poi}\left(\frac{\tilde{\mu}_i}{1+\rho}\right), \text{ for } t = 0, 1, \ldots, T,$$

where $\tilde{\mu}_i = \exp(\tilde{x}_i'\beta)$, $t = 0$ being an initial time. Next suppose that the response y_{it} is related to the d_{it} as

$$y_{it} = \rho * d_{i,t-1} + d_{it}, \text{ for } t = 1, \ldots, T, \tag{6.18}$$

where $\rho * d_{i,t-1} = \sum_{j=1}^{d_{i,t-1}} b_j(\rho)$ is the binomial thinning operation similar to (6.15). By similar calculations as in the AR(1) process, one obtains

$$\begin{aligned} E[Y_{it}] = \text{var}[Y_{it}] &= \tilde{\mu}_i \\ \text{corr}(Y_{it}, Y_{i,t-\ell}) &= c^*_{i,(t-\ell)t}(\rho) \\ &= \begin{cases} \rho/(1+\rho) & \text{for } \ell = 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{6.19}$$

Note that the lag correlations in (6.19) have the same forms as in the Gaussian MA(1) correlation structure, except that in the present set up $0 \leq \rho \leq 1$, whereas under the Gaussian structure $-1 < \rho < 1$.

6.3.3 *Poisson Equicorrelation Model*

Suppose that y_{i0} is a Poisson variable with the mean parameter $\tilde{\mu}_i = \exp(\tilde{x}_i'\beta)$. Also suppose that

$$d_{it} \overset{\text{iid}}{\sim} \text{Poi}(\tilde{\mu}_i(1-\rho)) \text{ for all } t = 1, \ldots, T.$$

By similar arguments as for the AR(1) and MA(1) processes, one can show that y_{it} given by

$$y_{it} = \rho * y_{i0} + d_{it} \tag{6.20}$$

also follows the Poisson distribution (i.e., $y_{it} \sim \text{Poi}(\tilde{\mu}_i)$, yielding the marginal properties

$$E[Y_{it}] = \text{var}[Y_{it}] = \tilde{\mu}_i = \exp(\tilde{x}_i'\beta). \tag{6.21}$$

Note that these marginal properties may also be computed directly by using the model (6.20). As far as the product moments properties are concerned, it can be shown that

$$\begin{aligned} \text{corr}(Y_{it}, Y_{i,t-\ell}) &= c^*_{i,(t-\ell)t}(\rho) \\ &= \rho, \end{aligned} \tag{6.22}$$

for all $\ell = 1, 2, \ldots, T-1$, with $0 \leq \rho \leq 1$ instead of $-(1/T-1) \leq \rho \leq 1$ under the Gaussian equicorrelation model.

For convenience, we summarize the means, variances, and correlations for all three stationary correlation models, as in Table 6.1.

Table 6.1 A class of stationary correlation models for longitudinal count data and basic properties.

Model	Dynamic Relationship	Mean, Variance, & Correlations
AR(1)	$y_{it} = \rho * y_{i,t-1} + d_{it}, t = 2, \ldots$	$E[Y_{it}] = \mu_{i\cdot}$
	$y_{i1} \sim Poi(\mu_{i\cdot})$	$\text{var}[Y_{it}] = \mu_{i\cdot}$
	$d_{it} \sim Poi(\mu_{i\cdot}(1-\rho)), t = 2, \ldots$	$\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho_\ell$
		$= \rho^\ell$
MA(1)	$y_{it} = \rho * d_{i,t-1} + d_{it}, t = 1, \ldots$	$E[Y_{it}] = \mu_{i\cdot}$
	$d_{i0} \sim Poi(\mu_{i\cdot}/(1+\rho))$	$\text{var}[Y_{it}] = \mu_{i\cdot}$
	$d_{it} \sim Poi(\mu_{i\cdot}/(1+\rho)), t = 1, \ldots$	$\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho_\ell$
		$= \begin{cases} \frac{\rho}{1+\rho} & \text{for } \ell = 1 \\ 0 & \text{otherwise,} \end{cases}$
EQC	$y_{it} = \rho * y_{i1} + d_{it}, t = 2, \ldots$	$E[Y_{it}] = \mu_{i\cdot}$
	$y_{i1} \sim Poi(\mu_{i\cdot})$	$\text{var}[Y_{it}] = \mu_{i\cdot}$
	$d_{it} \sim Poi(\mu_{i\cdot}(1-\rho)), t = 2, \ldots$	$\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho_\ell$
		$= \rho$

6.4 Inferences for Stationary Correlation Models

6.4.1 Likelihood Approach and Complexity

As opposed to the marginal likelihood estimation by (6.10), it is natural that under the correlation models (6.14), (6.18), and (6.20), the likelihood construction would be complicated. This is because under these models, the likelihood function is given by

$$L(\beta, \rho) = \Pi_{i=1}^K [f(y_{i1}) \Pi_{t=2}^T f(y_{it}|y_{i,t-1})], \tag{6.23}$$

where $f(y_{i1}) = \exp(-\tilde{\mu}_i)\tilde{\mu}_i^{y_{i1}}/y_{i1}!$ is the Poisson density with $\tilde{\mu}_i = \exp(\tilde{x}_i'\beta)$, under all three models, but the conditional densities $f(y_{it}|y_{i,t-1})$ would have different forms under different models. For example, under the stationary AR(1) model

(6.14), the conditional density has the form given by

$$f(y_{it}|y_{i,t-1}) = \sum_{s=1}^{\min(y_{it},y_{i,t-1})} \frac{(y_{i,t-1})!}{s!(y_{i,t-1}-s)!}\rho^s(1-\rho)^{y_{i,t-1}-s}\frac{\exp(-\tilde{\mu}_i)\tilde{\mu}_i^{y_{it}-s}}{(y_{it}-s)!}, \quad (6.24)$$

yielding by (6.23) a complex likelihood, which is not easy to maximize with regard to the desired parameters β and ρ.

In the following section we provide an alternative efficient approach for the estimation of the parameters of the models.

6.4.2 GQL Approach

Recall that under the independence assumption, one can solve the quasi-likelihood [QL; Wedderburn (1974)] estimating equation (6.9) for β, but this will be an inefficient estimate given that the repeated responses are now assumed to follow either the AR(1) correlation model (6.14) with correlation structure (6.17), MA(1) correlation model (6.18) with correlation structure (6.19), or equicorrelation model (6.20) with correlation structure as in (6.22). Note that all three correlation structures given in (6.17), (6.19), and (6.22), may be represented by a general autocorrelation matrix of the form

$$C_i^*(\rho) = (c^*_{i,(t-\ell)t}(\rho)) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix}, \quad (6.25)$$

[Sutradhar and Das (1999, Section 3)], where for $\ell = 1,\ldots,T-1$, ρ_ℓ represents the lag ℓ autocorrelation. For example, the AR(1) model based autocorrelation structure (6.17) may be represented by this correlation matrix $C_i^*(\rho)$ (6.25) by using $\rho_\ell = \rho^\ell$. Similarly, when one uses $\rho_1 = \rho/(1+\rho)$ and $\rho_2 = \rho_3 = \ldots = \rho_{T-1} = 0$, in (6.25), it produces the MA(1) correlation structure (6.19); and for $\rho_\ell = \rho$ for all $\ell = 1,\ldots,T-1$, $C_i^*(\rho)$ matrix in (6.25) represents the correlations under the equicorrelations structure (6.22).

It is therefore clear that if it is assumed that the repeated counted responses follow one of the AR(1), MA(1), or equi-correlation models, then one may estimate the regression effects under any of these three models by simply estimating this common $C_i^*(\rho)$ matrix in (6.25) and then using this estimated correlation matrix in a proper estimating equation for the regression effects β. Because $C_i^*(\rho)$ is the true correlation matrix for any of the three models, Sutradhar (2003, Section 3) proposed a generalized quasi-likelihood approach that generalizes the independence assumption based QL (6.9) approach of Wedderburn (1974) to the general stationary correlation setup. The GQL estimating equation for β is given by

$$\sum_{i=1}^{K} X_i' A_i \Sigma_i^{*-1}(\rho)(y_i - \mu_i) = 0, \tag{6.26}$$

where $\Sigma_i^*(\rho) = A_i^{1/2} C_i^*(\rho) A_i^{1/2}$, with $C_i^*(\rho)$ as the true stationary correlation structure for any of the AR(1), MA(1), or equicorrelation models. Note that in (6.26), $\mu_i = \tilde{\mu}_i \mathbf{1}$, $A_i = \text{diag}(\sigma_{itt}) = \text{diag}(\tilde{\mu}_i)$, as in (6.13), $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ is the $T \times 1$ vector of repeated counts for the ith individual, and $X_i' = [\tilde{x}_i, \ldots, \tilde{x}_i] : p \times T$ is the corresponding matrix of stationary covariates with $\tilde{x}_i = (x_{i1}, \ldots, x_{ip})'$ as the p-dimensional time-independent covariate vector as in (6.12).

Note that the GQL estimating equation (6.26) may be solved for β when ρ (i.e., all lag correlations $\rho_1, \ldots, \rho_\ell, \ldots, \rho_{T-1}$) is known. It is, however, not necessary to know the specific form for the correlation matrix $C_i^*(\rho)$, as this form in (6.25) is general which is valid under any of the three correlation structures (6.17), (6.19) and (6.22). In practice ρ is unknown, therefore the lag correlations can be consistently estimated by using the well-known method of moments. For $\ell = |u - t|$, $u \neq t$, $u, t = 1, \ldots, T$, the moment estimator for ρ_ℓ, the autocorrelation of lag ℓ, has the formula

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \tilde{y}_{it} \tilde{y}_{i,t+\ell} / K(T-\ell)}{\sum_{i=1}^{K} \sum_{t=1}^{T} \tilde{y}_{it}^2 / KT}, \tag{6.27}$$

[Sutradhar and Kovacevic (2000, eqn. (2.18)); Sutradhar (2003)], where $\tilde{y}_{it}$ is the standardized residual, defined as $\tilde{y}_{it} = (y_{it} - \mu_{it}) / \{\sigma_{itt}\}^{1/2}$. Note that under the present stationary correlation models for the repeated count data $\mu_{it} = \sigma_{itt} = \tilde{\mu}_i$ as in (6.12) and (6.13).

Let $\hat{\beta}_{GQL}$ denote the GQL estimator of β which is obtained by solving (6.26) after using $\hat{\rho}_\ell$ from (6.27) for ρ_ℓ. Note that because the left-hand side of the GQL estimating equation in (6.26) is an unbiased estimating function for the zero vector, $\hat{\beta}_{GQL}$, the root of the equation (6.26) is a consistent estimator for β.

6.4.2.1 Asymptotic Distribution of the GQL Estimator

Note that $\hat{\beta}_{GQL}$ may be obtained from (6.26) by using the iterative equation

$$\hat{\beta}_{GQL}(r+1) = \hat{\beta}_{GQL}(r) + \left[\sum_{i=1}^{K} X_i' \Sigma_i^{*-1}(\rho) X_i \right]_{(r)}^{-1} \times \left[\sum_{i=1}^{K} X_i' \Sigma_i^{*-1}(\rho)(y_i - \mu_i) \right]_{(r)}, \tag{6.28}$$

where $[\cdot]_{(r)}$ denotes that the expression within the brackets is evaluated at $\beta = \hat{\beta}_{GQL}(r)$, the rth iterative value for $\hat{\beta}_{GQL}$. Because $y_1, \ldots, y_i, \ldots, y_K$ are independent, by using the central limit theorem, it then follows from (6.28) that as $K \to \infty$, $(\hat{\beta}_{GQL} - \beta)$ has the p-dimensional multivariate normal distribution with mean vector

0 and $p \times p$ covariance matrix V^* given by

$$V^*(\hat{\beta}_{GQL}) = \lim_{K\to\infty}\left\{\sum_{i=1}^{K} X_i^T A_i^{1/2} C_i^{*-1}(\rho) A_i^{1/2} X_i\right\}^{-1}. \tag{6.29}$$

Note that this asymptotic distribution is given here for known ρ. This result, however, holds even when $\hat{\rho}$ is used for ρ. This is because it can be shown that $\hat{\rho}_\ell$ from (6.27) converges in probability to ρ_ℓ for all $\ell = 1,\ldots,T-1$.

6.4.2.2 'Working' Independence Assumption Based GQL Estimation

It is known that if one is interested in obtaining only a consistent estimator for β, this can be achieved by solving the GQL estimating equation (6.26) by pretending that the repeated responses are independent even though they are actually correlated following any of the three models (6.14), (6.18), or (6.20). Thus, we obtain a 'working' independence assumption based GQL estimate by solving

$$\sum_{i=1}^{K} X_i' A_i \Sigma_i^{*-1}(\rho)(y_i - \mu_i)|_{\rho=0} = \sum_{i=1}^{K} X_i'(y_i - \mu_i) = 0. \tag{6.30}$$

Note that this estimating equation is in fact the QL estimating equation (6.9) due to Wedderburn (1974), which is also the same as the MM estimating equation (6.5). This QL estimating equation is simpler to solve than the GQL (6.26) equation and this provides the consistent estimate for β.

Let $\hat{\beta}(I)$ denote the solution of (6.30). This estimator is the same as the MM estimator $\hat{\beta}_{MM}$ obtained from (6.5), therefore its asymptotic distribution is given by (6.7). Thus, $\hat{\beta}(I)$ has the asymptotic variance

$$V^*(\hat{\beta}(I)) = \text{limit}_{K\to\infty}\left[\sum_{i=1}^{K} X_i' A_i X\right]^{-1}\left[\sum_{i=1}^{K} X_i' \Sigma_i^*(\rho) X_i\right]\left[\sum_{i=1}^{K} X_i' A_i X\right]^{-1}, \tag{6.31}$$

where $\Sigma_i^*(\rho) = A_i^{1/2} C_i^*(\rho) A_i^{1/2}$.

6.4.2.3 Efficiency of the Independence Assumption Based Estimator

Similar to the correlated linear model case [Amemiya (1985, Section 6.1.3)], a comparison of (6.31) with (6.29) shows that the independence assumption based estimator $\hat{\beta}(I)$ always has the less than or the same efficiency asthe GQL estimator $\hat{\beta}_{GQL}$. We provide a numerical example below to illustrate this efficiency issue.

The percentage efficiency of the uth $(u = 1,\ldots,p)$ component of the $\hat{\beta}(I)$ estimator, for example, is defined as

$$\text{eff}(\hat{\beta}_u(I)) = \frac{\text{var}(\hat{\beta}_{u,GQL})}{\text{var}(\hat{\beta}_u(I))} \times 100, \tag{6.32}$$

where $\text{var}(\hat{\beta}_{u,GQL})$ and $\text{var}(\hat{\beta}_u(I))$ are the uth diagonal elements of the covariance matrices $V^*(\hat{\beta}_{GQL})$ (6.29) and $V^*(\hat{\beta}(I))$ (6.31), respectively. Let us take $p = 2$ for simplicity so that the Poisson mean and the variance μ_{it} for the ith $(i = 1,\ldots,K)$ at time t $(t = 1,\ldots,T)$, has the formula $\tilde{\mu}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)$ under any of the three stationary models (6.14), (6.18), or (6.20). Let us consider $K = 100$, and three values of $T = 5, 10,$ and 15. As far as the time-independent stationary covariates are considered, we choose

$$x_{it1} = \tilde{x}_{i1} = 1.0, \text{ for all } i = 1,\ldots,K, \text{ and } t = 1,\ldots,T,$$

and

$$x_{it2} = \tilde{x}_{i2} = \begin{cases} -1 & \text{for } t = 1,\ldots,T;\ i = 1,\ldots,K/4 \\ 0 & \text{for } t = 1,\ldots,T;\ i = (K/4)+1,\ldots,K/2 \\ 0 & \text{for } t = 1,\ldots,T;\ i = (K/2)+1,\ldots,3K/4 \\ 1 & \text{for } t = 1,\ldots,T;\ i = (3K/4)+1,\ldots,K; \end{cases}$$

Next to compute the covariance matrices $V^*(\hat{\beta}_{GQL})$ (6.29) and $V^*(\hat{\beta}(I))$ (6.31), we need to construct the X_i and A_i matrices by

$$X_i = [\tilde{x}_{i1}\mathbf{1}_T, \tilde{x}_{i2}\mathbf{1}_T], \text{ and } A_i = \text{diag}[\tilde{\mu}_i] \ : T \times T.$$

We also need to specify the correlation matrix $C_i^*(\rho)$. We choose all three correlation models AR(1), MA(1), and exchangeable correlation structures given by (6.17), (6.19), and (6.22), respectively. Note that because the lag 1 correlations under the AR(1) (6.17) and equicorrelations (6.22) structures are given as $\rho_1 = \rho$, we choose, for example, $\rho = 0.3$ and 0.7 under both AR(1) and equi-correlation structures. But, as the lag 1 correlation under the MA(1) structure has to satisfy the range $0 < \rho_1 = \rho/(1+\rho) < 0.5$, we choose, for example, two values of $\rho = 0.25$ and 0.67, yielding the lag 1 correlations $\rho_1 = 0.2$ and 0.4, respectively.

For $\beta_1 = \beta_2 = 1.0$, and for the selected values of ρ, the efficiencies of $\hat{\beta}(I)$ as compared to $\hat{\beta}_{GQL}$ are given in Table 1.

The results of Table 6.2 show that as expected the independence assumption based GQL estimator $\hat{\beta}(I)$ obtained by solving (6.30) always has less or the same efficiency as compared to the true correlation structure based GQL estimator $\hat{\beta}_{GQL}$ obtained by solving (6.26).

Table 6.2 Percentage relative efficiency of $\hat{\beta}_1(I)$ and $\hat{\beta}_2(I)$ to the generalized estimators $\hat{\beta}_{1,GQL}$ and $\hat{\beta}_{2,GQL}$, respectively, with true stationary correlation matrix $C_1^*(\rho)$ for AR(1), MA(1), and Equi-correlation structures, for $\mu_{it} = \tilde{\mu}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)$ with $\beta_1 = \beta_2 = 1$

	AR(1)			MA(1)			EQC		
T	ρ	$\hat{\beta}_1(I)$	$\hat{\beta}_2(I)$	ρ	$\hat{\beta}_1(I)$	$\hat{\beta}_2(I)$	ρ	$\hat{\beta}_1(I)$	$\hat{\beta}_2(I)$
5	0.3	98	98	0.25	99	99	0.30	100	100
	0.49	96	96				0.49	100	100
	0.7	95	95	0.67	97	97	0.7	100	100
10	0.3	99	99	0.25	99	99	0.3	100	100
	0.49	96	96				0.49	100	100
	0.7	93	93	0.67	98	98	0.7	100	100
15	0.3	99	99	0.25	100	100	0.3	100	100
	0.49	97	97				0.49	100	100
	0.7	93	93	0.67	99	99	0.7	100	100

6.4.2.4 Performance of the GQL Estimation: A Simulation Example

Suppose that the repeated count responses follow either of the three stationary, namely AR(1)(6.17), MA(1) (6.19), or equicorrelation (6.22) structures. In estimating the regression effects β, the GQL approach does not, however, require us to know the specific correlation structure. What is needed here is: first consider that the repeated data for the ith individual has the autocorrelation matrix $C_i^*(\rho)$ (6.25) which in fact is a valid matrix not only for the above three correlation structures but also for any higher-order such as AR(2) and MA(2) correlation structures. Second, estimate this general autocorrelation matrix consistently and use the estimate in the GQL estimating equation (6.26) for β. This prompts the following two-step estimation.

Step 1. First, we solve the estimating equation for β (6.26) iteratively by (6.28), using starting values zero for longitudinal correlations and small positive or negative values for the regression parameters.
Step 2. This interim estimate of β from step 1 is then used in (6.27) to obtain the estimate of the autocorrelation matrix $C_i^*(\rho)$ in (6.25), which is used in turn in (6.28) to compute the new β estimate. This cycle of iterations continues until convergence.

To examine the performance of the above two-step based GQL estimation, we now consider a simulation study. Suppose that we follow the Poisson AR(1) model (6.14) and generate $T = 4$ repeated count observations for each of $K = 100$ independent individuals. As far as the covariates are concerned, we choose $p = 2$ time-independent covariates for each of these 100 individuals, given by

$$x_{it1} = \begin{cases} -1 & \text{for } t = 1,\ldots,T;\ i = 1,\ldots,K/4 \\ 0 & \text{for } t = 1,\ldots,T;\ i = (K/4)+1,\ldots,K/2 \\ 0 & \text{for } t = 1,\ldots,T;\ i = (K/2)+1,\ldots,3K/4 \\ 1 & \text{for } t = 1,\ldots,T;\ i = (3K/4)+1,\ldots,K; \end{cases}$$

and

$$x_{it2} = z_i^* mbox for\, t = 1,\ldots,T;\ i = 1,\ldots,K,$$

where z_i^* is a standard normal quantity. In this problem, $\beta = (\beta_1, \beta_2)'$ denotes the effects of the two covariates on the repeated counts.

Note that even though the data are generated following the AR(1) model (6.14), the GQL approach does not, however, require this model to be known for the estimation of β. This is because the GQL estimating equation (6.26) is developed based on a general autocorrelation structure $C_i(\rho^*)$, which accommodates all three AR(1) (6.17), MA(1) (6.19), and exchangeable (6.22) correlation structures. Further note that for $T = 4$, this general autocorrelation structure has three lag correlations, namely, ρ_1, ρ_2, and ρ_3, to estimate, by using the formula (6.27) as explained in Step 2 above. It would be interesting to see how these three estimates behave in estimating the three lag correlations ρ, ρ^2, and ρ^3, for the AR(1) model that generated the data. Next these correlation estimates are used in step 1 to estimate β by solving the GQL estimating equation (6.26). For a selected set of parameter values, namely $\beta_1 = \beta_2 = 0.0$, and $\rho = 0.6,\ 0.8$, the simulation is repeated 500 times. The average and standard error of the 500 estimates for each parameter are given in Table 6.3. In the table, these estimates are referred to as the simulated mean (SM) and simulated standard error (SSE). The estimated standard errors (ESE) of the regression estimates are also computed. This is done by using the asymptotic covariance formula for $V^*(\hat{\beta}_{GQL})$ given in (6.29).

Table 6.3 Simulated means, simulated standard errors, and estimated standard errors of the GQL estimates for regression and autocorrelation coefficients for selected values of the true correlation parameter under the Poisson AR(1) process with $T = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, based on 500 simulations.

		Estimates				
AR(1) Correlation (ρ)	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.6	SM	-0.003	-0.001	0.595	0.352	0.203
	SSE	0.085	0.049	0.061	0.088	0.108
	ESE	0.086	0.050			
0.8	SM	0.000	0.003	0.791	0.626	0.496
	SSE	0.096	0.056	0.043	0.070	0.098
	ESE	0.098	0.057			

The results in Table 6.3 clearly show that the two-step based GQL approach estimates all parameters very well. For example, when $\rho = 0.8$, the lag correlation estimates are 0.791, 0.626, and 0.496, whereas the true AR(1) based lag correlations are $\rho = 0.8$, $\rho^2 = 0.64$, and $\rho^3 = 0.512$. Similarly, the GQL approach estimates for $\beta_1 = \beta_2 = 0$ are 0.000, 0.003. Furthermore, for this $\rho = 0.8$ case, the ESE of the regression estimates , that is, 0.098, and 0.0.57 appear to be very close to the SSEs 0.096 and 0.056, respectively.

In Tables 6.4 and 6.5 below, we show similar results with regard to the performance of the GQL approach when data are generated under the MA(1) (6.18) and exchangeable (6.20) correlation models, respectively, by using the same covariates as in the AR(1) case.

Table 6.4 Simulated means, simulated standard errors, and estimated standard errors of the GQL estimates for regression and autocorrelation coefficients for selected values of the true correlation parameter under the Poisson MA(1) process with $T = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, based on 500 simulations.

		Estimates				
ρ (MA(1) Correlation (ρ_1))	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.25 (0.2)	SM	0.002	0.002	0.191	-0.006	0.004
	SSE	0.083	0.063	0.058	0.073	0.100
	ESE	0.081	0.063			
0.67 (0.4)	SM	-0.004	-0.004	0.396	-0.005	-0.004
	SSE	0.085	0.069	0.059	0.074	0.097
	ESE	0.088	0.070			

Table 6.5 Simulated means, simulated standard errors, and estimated standard errors of the GQL estimates for regression and autocorrelation coefficients for selected values of the true correlation parameter under the Poisson equicorrelation process with $T = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, based on 500 simulations.

		Estimates				
Equi-correlation (ρ)	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.6	SM	-0.006	-0.005	0.587	0.587	0.587
	SSE	0.119	0.096	0.064	0.065	0.088
	ESE	0.118	0.093			
0.8	SM	-0.009	-0.009	0.790	0.790	0.789
	SSE	0.131	0.101	0.043	0.041	0.059
	ESE	0.130	0.103			

6.4.3 GEE Approach and Limitations

In order to gain efficiency over the independence assumption based regression estimator $\hat{\beta}(I)$ (6.30), in the generalized estimating equations approach [Liang and Zeger (1986)], one solves a 'working' correlation matrix, $R(\alpha)$, based estimating equation

$$\sum_{i=1}^{K} X_i' A_i {V^*}_i^{-1}(\hat{\alpha})(y_i - \mu_i) = 0, \tag{6.33}$$

where $V_i^*(\alpha) = A_i^{1/2} R(\alpha) A_i^{1/2}$ is the working covariance matrix of y_i, α being an $s \times 1$ vector of parameters which fully characterizes $R(\alpha)$. Note that the GEE in (6.33) appears to be similar to the GQL estimating equations in (6.26), but they are quite different. Also, in (6.33), $\hat{\alpha}$ is obtained by solving a 'working' correlation model based moment equation. The data used in such a moment equation follow a different but true correlation structure, thus it is inappropriate to assume that $\hat{\alpha}$ converges to α [Crowder (1995)]. In view of this anomaly, any efficiency computations by using $\hat{\alpha}$ for α in the formula for the covariance matrix of the GEE estimator obtained from (6.33) [Liang and Zeger (1986)] would be incorrect.

Let $\hat{\beta}_G$ be the solution for β based on (6.33). Next suppose that $\hat{\alpha}$ converges to α_0, which must be a function of the true correlation parameter (ρ). In order to examine the correlation misspecification effects on the efficiency of $\hat{\beta}_G$, Sutradhar and Das (1999) have suggested using this α_0 in the formula for the covariance matrix of $\hat{\beta}_G$. Thus, $K^{1/2}(\hat{\beta}_G - \beta)$ is now asymptotically multivariate Gaussian with zero mean vector and covariance matrix V_G given by

$$\begin{aligned} V_G = \lim_{K\to\infty} K \left(\sum_{i=1}^{K} X_i' A_i^{1/2} R^{-1}(\alpha_0) A_i^{1/2} X_i \right)^{-1} \\ \times \left\{ \sum_{i=1}^{K} X_i' A_i^{1/2} R^{-1}(\alpha_0) C_i(\rho) R^{-1}(\alpha_0) A_i^{1/2} X_i \right\} \\ \times \left\{ \sum_{i=1}^{K} X_i' A_i^{1/2} R^{-1}(\alpha_0) A_i^{1/2} X_i \right\}^{-1}, \end{aligned} \tag{6.34}$$

where $C_i(\rho)$ is the true correlation matrix, as given in (6.25).

6.4.3.1 Efficiency of the GEE Based Estimator Under Correlation Structure Mis-specification

As far as the correlation models are concerned, we consider the same three stationary Poisson correlation models as we took for Section 6.4.2.3. Note that similar to (6.32), the percentage efficiency of the uth $(u = 1, \ldots, p)$ component of the $\hat{\beta}_G$

estimator, for example, is defined as

$$\text{eff}(\hat{\beta}_{u,G}) = \frac{\text{var}(\hat{\beta}_{u,TR})}{\text{var}(\hat{\beta}_{u,G})} \times 100, \tag{6.35}$$

where $\text{var}(\hat{\beta}_{u,TR})$ is the uth diagonal element of the covariance matrix of the true correlation structure based estimator V^*_{TR} computed by (6.29) using the true correlation structure for $C^*_i(\rho)$, and $\text{var}(\hat{\beta}_{u,G})$ is the uth diagonal element of the covariance matrix V_G given in (6.34). For the purpose, we first show how to compute α_0 under possible model mis-specifications, and then compute the efficiencies.

(i) Computation of α_0 Under True AR(1) Correlation Structure For EQC Working Correlation Structure

Under the working exchangeable correlation structure, $\hat{\alpha}$ satisfies the estimating equation

$$\sum_{i=1}^{K}\sum_{t\neq u}^{T}(\tilde{y}_{it}\tilde{y}_{iu} - \alpha) = 0, \tag{6.36}$$

where $\tilde{y}_{it} = (y_{it} - \mu_{it})/\{\sigma_{itt}\}^{1/2}$, as in (6.27), with $\mu_{it} = \sigma_{itt} = \tilde{\mu}_i = \exp(\tilde{x}'_i\beta)$ for the present stationary case. Note that for the true AR(1) correlation structure, $E(\tilde{y}_{it}\tilde{y}_{iu}) = \rho^{|t-u|}$ with $0 < \rho < 1$. This shows that $\hat{\alpha}$ obtained from (6.36), if it exists, will converge to α_0 satisfying

$$\alpha_0 = 2\rho\{T - (1-\rho^T)/(1-\rho)\}/T(T-1)(1-\rho). \tag{6.37}$$

For example, when $\rho = 0.7$ the equation (6.37) yields $\alpha_0 = 0.52$, 0.35 and 0.26 for $T = 5$, 10, and 15, respectively.

Now to compute the efficiency of the 'working' equicorrelation structure based GEE estimator $\hat{\beta}_G$, when in fact the repeated counts truly follow the AR(1) correlation structure, we need to put AR(1) based $C_i(\rho)$ and EQC based $R(\alpha_0)$ in (6.34), for example, with $\alpha_0 = 0.52$ when $\rho = 0.7$ for $T = 5$. The efficiencies for selected ρ and for the selected design covariates as in Section 6.4.2.4 for $T = 5,\ 10,\ 15$, are shown in Table 6.6.

For MA(1) Working Correlation Structure

For the working MA(1) correlation structure, we solve

$$\sum_{i=1}^{K}\sum_{t=1}^{T-1}\tilde{y}_{it}\tilde{y}_{i(t+1)} - K(T-1)\alpha = 0, \tag{6.38}$$

to obtain $\hat{\alpha}$. If $\hat{\alpha}$ exists, then in this case $\hat{\alpha}$ will converge to $\alpha_0 = \rho$, because, under the true AR(1) structure, $E(\tilde{y}_{it}\tilde{y}_{i(t+1)}) = \rho$. Note, however, that although in the present case ρ can take any value from 0 to 1, we can use only the range

Table 6.6 Percentage relative efficiency of $\hat{\beta}_{1,G}$ and $\hat{\beta}_{2,G}$ to the true correlation structure based estimators $\hat{\beta}_{1,T}(=\hat{\beta}_{1,GQL})$ and $\hat{\beta}_{2,T}(=\hat{\beta}_{2,GQL})$, respectively, with true stationary correlation matrix $C_1^*(\rho)$ for AR(1) structure, for $\mu_{it}=\tilde{\mu}_i=\exp(\tilde{x}_{i1}\beta_1+\tilde{x}_{i2}\beta_2)$ with $\beta_1=\beta_2=1$

True Correlation Structure AR(1)

Working Structure	MA(1)				EQC			
T	ρ	α_0	$\hat{\beta}_{1,MA(1)}$	$\hat{\beta}_{2,MA(1)}$	ρ	α_0	$\hat{\beta}_{1,EQC}$	$\hat{\beta}_{2,EQC}$
5	0.3	0.3	100	100	0.3	0.15	98	98
	0.49	0.49	95	95	0.7	0.52	95	95
10	0.3	0.3	100	100	0.3	0.08	99	99
	0.49	0.49	98	98	0.7	0.35	93	93
15	0.3	0.3	100	100	0.3	0.06	99	99
	0.49	0.49	97	97	0.7	0.26	93	93

$0<\rho(=\alpha_0)<0.5$ for the efficiency computation. This is because in the GEE approach ρ is unknown and the working correlation α can range from -0.5 to 0.5 only. This is clear from the formula of V_G in (6.34), where one cannot use $R^{-1}(\alpha_0)$ beyond the range $-0.5<\alpha<0.5$, as $R(\alpha)$ has the MA(1) correlation structure. In view of this we have chosen $\rho=0.3$ and 0.49 for our efficiency computations. These efficiencies are also reported in Table 6.6, for $T=5,\ 10$, and 15.

(ii) Computation of α_0 Under True MA(1) Correlation Structure For AR(1) Working Correlation Structure

Let $c_{i,ut}$ be the (u,t) element of the true correlation matrix $C_i(\rho)$. For MA(1) true correlation structure, $c_{i,ut}=\rho_1=\rho(1+\rho)$ if $|t-u|=1$, and $c_{i,ut}=0$ otherwise, where ρ_1 denotes the lag-1 correlation. Under this structure, ρ_1 satisfies $-0.5\leq\rho_1\leq 0.5$.

Now consider the working AR(1) correlation matrix. Here $r_{i,ut}=\alpha^{|t-u|}$ for $u,t=1,\ldots,T$. If we base the estimation again on the average correlation, the estimating equation

$$\sum_{i=1}^{K}\sum_{u<t}^{T}(\tilde{y}_{it}\tilde{y}_{iu}-\alpha^{|t-u|})=0 \tag{6.39}$$

results, giving $\hat{\alpha}$; a simple moment estimator for α, see also Crowder (1995), where $\tilde{y}_{iu}$ and $\tilde{y}_{it}$ are the standardized residuals defined as in (6.36). Because

$$E\left\{\sum_{u<t}^{T}\tilde{y}_{it}\tilde{y}_{iu}\right\}=(T-1)\rho_1=(T-1)\frac{\rho}{1+\rho}$$

under the MA(1) correlation structure, it follows from (6.39) that α_0 is in fact the solution of

$$\alpha_0(1-\alpha_0)^{-1}\{T-(1-\alpha_0^T)/(1-\alpha_0)\}-(T-1)\rho_1=0. \tag{6.40}$$

Therefore, if $\hat{\alpha}$ exists, $\hat{\alpha}$ will converge in probability to α_0, α_0 being related to ρ through (6.40). For example, when $\rho_1 = 0.4$, that is, $\rho = 0.67$, the α_0 values are approximately 0.31, 0.30, and 0.29 for $T = 5$, 10, and 15 respectively. For selected values of ρ, the efficiencies of $\hat{\beta}_G$ for the MA(1) versus AR(1) correlation structures, are shown in Table 6.7.

Table 6.7 Percentage relative efficiency of $\hat{\beta}_{1,G}$ and $\hat{\beta}_{2,G}$ to the true correlation structure based estimators $\hat{\beta}_{1,TR}(=\hat{\beta}_{1,GQL})$ and $\hat{\beta}_{2,TR}(=\hat{\beta}_{2,GQL})$, respectively, with true stationary correlation matrix $C_1^*(\rho)$ for MA(1) structure, for $\mu_{it} = \tilde{\mu}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)$ with $\beta_1 = \beta_2 = 1$

True Correlation Structure MA(1)

Working Structure	AR(1)				EQC			
T	ρ	α_0	$\hat{\beta}_{1,AR(1)}$	$\hat{\beta}_{2,AR(1)}$	ρ	α_0	$\hat{\beta}_{1,EQC}$	$\hat{\beta}_{2,EQC}$
5	0.25	0.17	100	100	0.25	0.08	99	99
	0.67	0.31	99	99	0.67	0.16	97	97
10	0.25	0.17	100	100	0.25	0.04	99	99
	0.67	0.30	100	100	0.67	0.08	98	98
15	0.25	0.17	100	100	0.25	0.04	99	99
	0.67	0.29	100	100	0.67	0.05	98	98

For EQC Working Correlation Structure

For the working exchangeable correlation matrix $R(\alpha)$, one writes $r_{i,ut} = \alpha$ for all u, t except for $u = t$. We must have $-\{1/(T-1)\} \leq \alpha \leq 1$ for $R(\alpha)$ to be a positive definite matrix, where T is the dimension of the $R(\alpha)$ matrix. It then follows that the moment estimator $\hat{\alpha}$ [see also Crowder (1995] for α is given by

$$\hat{\alpha} = \sum_{i=1}^{K} \sum_{u \neq t}^{T} \hat{r}_{i(ut)} / KT(T-1)$$

$$= \sum_{i=1}^{K} \sum_{u \neq t}^{T} \tilde{y}_{iu}\tilde{y}_{it} / KT(T-1). \tag{6.41}$$

Because $C_i(\rho)$ has the MA(1) correlation structure,

$$E(\hat{\alpha}) = \{KT(T-1)\}^{-1} 2K(T-1)\rho_1 = 2\rho_1/T = \frac{2\rho}{T(1+\rho)}. \tag{6.42}$$

Thus, if $\hat{\alpha}$ exists, then $\hat{\alpha}$ converges to $\alpha_0 = 2\rho_1/T$. Therefore, to compute the efficiency of $\hat{\beta}_G$, we use the true $\rho_1 = \rho/(1+\rho)$ for $C_i(\rho)$ and $\alpha_0 = 2\rho_1/T$ for $R(\alpha_0)$ in V_G given in (6.34). For example, with $T = 5$ and $\rho = 0.67$, we use $\alpha_0 = 0.16$ in $R(\alpha_0)$. The efficiencies for selected values of ρ are shown in Table 6.7.

(iii) Computation of α_0 Under True Equicorrelation (EQC) Structure

For AR(1) Working Correlation Structure:

For the working AR(1) correlation structure, the estimating equation for α remains the same as (6.39). However, as $E(\tilde{y}_{iu}\tilde{y}_{it}) = \rho$ under the true exchangeable correlation structure, $\hat{\alpha}$ obtained from (6.39), if it exists, converges to α_0, now satisfying the equation

$$\alpha_0(1-\alpha_0)^{-1}\{T-(1-\alpha_0^T)/(1-\alpha_0)\} - T(T-1)\rho/2 = 0. \quad (6.43)$$

Here $\rho \geq -1/(T-1)$. Consequently, we use only positive ρ values for efficiency computations. For example, when $\rho = 0.7$ is used in (6.43), α_0 is 0.83, 0.90, and 0.93 for $T = 5$, 10, and 15 respectively. Now the efficiencies of AR(1) 'working' structure based $\hat{\beta}_G$, when EQC is the true correlation structure, are shown in Table 6.8, for the selected values of ρ.

Table 6.8 Percentage relative efficiency of $\hat{\beta}_{1,G}$ and $\hat{\beta}_{2,G}$ to the true correlation structure based estimators $\hat{\beta}_{1,TR}(=\hat{\beta}_{1,GQL})$ and $\hat{\beta}_{2,TR}(=\hat{\beta}_{2,GQL})$, respectively, with true stationary correlation matrix $C_1^*(\rho)$ for EQC structure, for $\mu_{it} = \tilde{\mu}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)$ with $\beta_1 = \beta_2 = 1$.

EQC True Correlation Structure

Working Structure	AR(1)				MA(1)			
T	ρ	α_0	$\hat{\beta}_{1,AR(1)}$	$\hat{\beta}_{2,AR(1)}$	ρ	α_0	$\hat{\beta}_{1,MA(1)}$	$\hat{\beta}_{2,MA(1)}$
5	0.3	0.49	96	96	0.3	0.3	99	99
	0.7	0.83	95	95	0.49	0.49	92	92
10	0.3	0.65	95	95	0.3	0.3	99	99
	0.7	0.90	94	94	0.49	0.49	98	98
15	0.3	0.74	94	94	0.3	0.3	100	100
	0.7	0.93	93	93	0.49	0.49	98	98

For MA(1) Working Correlation Structure

For the working MA(1) correlation structure, the estimating equation for α is given by (6.38). Because $E(\tilde{y}_{it}\tilde{y}_{i(t+1)}) = \rho$ for the true exchangeable correlation structure, it follows from (6.38) that $\hat{\alpha}$, if it exists, converges to $\alpha_0 = \rho$. The efficiencies of $\hat{\beta}_G$ for the exchangeable versus MA(1) correlation structure are also shown in Table 6.8, for selected values of ρ.

Note that when the efficiencies displayed in Tables 6.6 – 6.8 under correlation structure misspecification are compared with those in Table 6.2 computed for the independence assumption based regression estimators, it is seen that in some cases, especially when EQC is the true correlation structure, the $\hat{\beta}(I)$ appears to be equally or more efficient than the GEE based estimator $\hat{\beta}_G$. For this reason, as Sutradhar and Das (1999) [see also Sutradhar (2003)] argued, there is no guarantee that the

GEE approach can provide more efficient estimates than the simpler MM estimates obtained from (6.6) or QL estimates obtained from (6.9).

6.5 Nonstationary Correlation Models

In Section 6.3, we provided three stationary correlation models for longitudinal count data. In Section 6.4, we discussed various estimation techniques including the GEE and GQL approaches, for the estimation of the regression effects. Note that in the GEE approach, the selection of a suitable 'working' correlation structure out of these three or other possible correlation structures is left to the user. It was shown in Section 6.4 [see also Sutradhar and Das (1999)] that the use of such a 'working' correlation structure may in reality produce a less efficient estimate for the regression effect β than the 'independence' assumption based estimate. As a remedy, Sutradhar (2003) has suggested using a general (robust) autocorrelation structure that accommodates the above three stationary correlation structures as special cases. Thus, as demonstrated in Section 6.4.2.3 (see Table 6.2), if the data follow this class of Gaussian type stationary correlation structure, then the solution of a generalized quasi-likelihood equation, following Sutradhar (2003), always produces consistent and efficient estimates.

There, however, remains a concern that it may not be reasonable to use a stationary correlation structure when it is known that the covariates are time dependent. In Section 6.5.1, we provide three nonstationary correlation models as a generalization of the stationary AR(1), MA(1), and EQC structures, discussed in Section 6.3. These models produce the same mean and variance functions, and different correlation structures, under both stationary and nonstationary conditions. Under the assumption that the repeated count data follow one of these three possible nonstationary models, in Section 6.5.2, we discuss the estimation of the parameters under all three models. In Section 6.6.1, we deal with more nonstationary autocorrelation models that belong to the same autocorrelation class as that of Section 6.5, but now the marginal means and variances can be different under different models. In Section 6.6.2 we provide a model selection criterion based on the principle of minimum error sum of squares. A simulation study is conducted in Section 6.6.3 to examine the performances of the estimates under the true as well as misspecified models. Also, the simulation study in the same section justifies the model selection criterion. In Section 6.7, a real-life data example is discussed both for model selection as well as estimation of the regression effects and the correlation parameters.

6.5.1 Nonstationary Correlation Models with the Same Specified Marginal Mean and Variance Functions

6.5.1.1 Nonstationary AR(1) Models

Suppose that y_{i1} follows the Poisson distribution with mean parameter $\mu_{i1} = \exp(x'_{i1}\beta)$; that is, $y_{i1} \sim \text{Poi}(\mu_{i1} = \exp(x'_{i1}\beta))$, and for $t = 2,\ldots,T$, y_{it} relates to $y_{i,t-1}$ through the dynamic relationship

$$y_{it} = \rho * y_{i,t-1} + d_{it}, \text{ for } t = 2,\ldots,T, \tag{6.44}$$

where

$$\rho * y_{i,t-1} = \sum_{s=1}^{y_{i,t-1}} b_s(\rho),$$

with $Pr[b_s(\rho) = 1] = \rho$ and $Pr[b_s(\rho) = 0] = 1-\rho$. Also suppose that

$$y_{i,t-1} \sim \text{Poi}(\mu_{i,t-1}), \text{ and } d_{it} \sim \text{Poi}(\mu_{it} - \rho\mu_{i,t-1}),$$

with $\mu_{it} = e^{x'_{it}\beta}$, and d_{it} and $y_{i,t-1}$ are independent. After some algebra, it may be shown that this model (6.44) yields the means and the variances as

$$E(Y_{it}) = \text{var}(Y_{it}) = \mu_{it} = e^{x'_{it}\beta}, \tag{6.45}$$

and for $u < t$ with $t = 2,\ldots,T$, nonstationary (ns) correlations, say $c^{(ns)}_{i,ut}(x_{iu}, x_{it}, \rho)$, as

$$\begin{aligned} corr(Y_{iu}, Y_{it}) &= c^{(ns)}_{i,ut}(x_{iu}, x_{it}, \rho) \\ &= \rho^{t-u}\sqrt{\frac{\mu_{iu}}{\mu_{it}}}, \end{aligned} \tag{6.46}$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \frac{\mu_{it}}{\mu_{i,t-1}}\right], \ t = 2,\cdots,T. \tag{6.47}$$

Stationary Correlation Structure: Note that in the stationary case, that is, when the covariates are time independent such as $x_{it} = \tilde{x}_i$ for all $t = 1,\ldots,T$, the means and variances given by (6.45) and the correlation matrix given by (6.46) become stationary. In particular, the nonstationary correlations given by (6.46) reduce to the covariates free stationary correlations

$$c^*_{i,ut}(\rho)) = (\rho^{|t-u|}), \text{ for all } u \neq t, u,t = 1,\ldots,T, \tag{6.48}$$

which is same as the correlation in (6.17) derived under the stationary correlation model (6.14).

6.5.1.2 Nonstationary MA(1) Models

To generalize the stationary MA(1) model [Sutradhar (2003)] to the nonstationarity case, we consider the dynamic relationship

$$\begin{aligned} y_{i1} &\sim \text{Poi}(\mu_{i1} = \exp(x'_{i1}\beta)) \\ y_{it} &= \rho * d_{i,t-1} + d_{it}, \ \text{for } t = 2, \ldots, T, \end{aligned} \tag{6.49}$$

where

$$d_{it} \overset{iid}{\sim} \text{Poi}\left[\sum_{j=0}^{t-1} (-\rho)^j \mu_{i,t-j}\right] \ \text{for all } t = 1, \ldots, T.$$

After some algebra, this model yields the same means and variances as in (6.45) derived under the AR(1) model. Furthermore, it can be shown that the correlations are given by

$$\text{corr}(Y_{iu}, Y_{it}) = c^{(ns)}_{i,ut}(x_{iu}, x_{it}, \rho) = \begin{cases} \frac{\rho\{\sum_{j=0}^{\min(u,t)-1} (-\rho)^j \mu_{i,\min(u,t)-j}\}}{\sqrt{\mu_{iu}\mu_{it}}} & \text{for } |u-t| = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6.50}$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \rho_{i20}, \ldots, \rho_{it0}, \ldots, \rho_{iT0}\right], \tag{6.51}$$

where ρ_{it0} is the solution of $\sum_{j=0}^{t-1}(-\rho)^j \mu_{i,t-j} = 0$. Note that this range restriction may allow only a narrow range for the ρ parameter.

Stationary Correlation Structure: Note that in the stationary case, the means and the variances have the form $\mu_{it} = \mu_{i\cdot} = \exp(\tilde{x}'_i\beta)$ for all $t = 1, \ldots, T$. Furthermore, by (6.50), the limiting correlations when $\min(u,t) \to \infty$ have the formula

$$c^*_{i,ut}(\rho) = \text{corr}(Y_{iu}, Y_{it}) = \begin{cases} \rho\{\sum_{j=0}^{\infty}(-\rho)^j = \frac{\rho}{1+\rho} & \text{for} |u-t| = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6.52}$$

which is free from the time-dependent covariates. This stationary correlation is the same as the correlation in (6.19) derived under the stationary MA(1) model (6.18).

6.5.1.3 Nonstationary EQC Models

To generate a nonstationary equicorrelations model, we consider

$$\begin{aligned} y_{i1} &\sim \text{Poi}(\mu_{i1} = \exp(x'_{i1}\beta)) \\ y_{it} &= \rho * y_{i1} + d_{it}, \ \text{for } t = 2, \ldots, T, \end{aligned} \tag{6.53}$$

where d_{it} is assumed to be distributed as

$$d_{it} \sim \text{Poi}(\mu_{it} - \rho\mu_{i1})$$

with $\mu_{it} = e^{x'_{it}\beta}$. Also it is assumed that d_{it} for $t = 2,\ldots,T$, are independent of y_{i1}. It then follows that $E(Y_{it}) = \text{var}(Y_{it}) = \mu_{it} = e^{x'_{it}\beta}$ as in the AR(1) and MA(1) cases, for all $t = 1,\ldots,T$, and for $u < t$,

$$\text{cov}(Y_{iu}, Y_{it}) = \rho\mu_{i1}, \tag{6.54}$$

yielding the nonstationary correlation structure

$$\text{corr}(Y_{iu}, Y_{it}) = c^{(ns)}_{i,ut}(x_{iu}, x_{it}, \rho) = \frac{\rho\mu_{i1}}{\sqrt{\mu_{iu}\mu_{it}}}, \tag{6.55}$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \frac{\mu_{it}}{\mu_{i1}}\right], \ t = 2,\ldots,T.$$

Stationary Correlation Structure: Note that when covariates are time independent, that is, $x_{it} = \tilde{x}_i$ for all $t = 1,\ldots,T$, the nonstationary correlations in (6.55) reduce to the stationary correlations in (6.22) derived under the stationary exchangeable correlation model (6.20).

For convenience, we summarize the means, variances, and correlations for all three nonstationary correlation models, as in Table 6.9.

Table 6.9 A class of nonstationary correlation models for longitudinal count data and basic properties.

Model	Dynamic Relationship	Mean, Variance and Correlations
AR(1)	$y_{it} = \rho * y_{i,t-1} + d_{it}, t = 2,\ldots,T$ $y_{i1} \sim Poi(\mu_{i1})$ $d_{it} \sim Poi(\mu_{it} - \rho\mu_{i,t-1}), t = 2,\ldots,T$	$E[Y_{it}] = \mu_{it}$ $\text{var}[Y_{it}] = \mu_{it}$ $\text{corr}[Y_{iu}, Y_{it}] = \rho^{(ns)}_{\lvert t-u\rvert}$ $= \rho^{\lvert t-u\rvert}\left[\frac{\mu_{iu}}{\mu_{it}}\right]^{\frac{1}{2}}$
MA(1)	$y_{it} = \rho * d_{i,t-1} + d_{it}, t = 2,\ldots,T$ $y_{i1} \sim Poi(\mu_{i1})$ $d_{it} \overset{iid}{\sim} Poi\left[\Sigma_{j=0}^{t-1}(-\rho)^j\mu_{i,t-j}\right] \ t = 1,\ldots,T$	$E[Y_{it}] = \mu_{it}$ $\text{var}[Y_{it}] = \mu_{it}$ $\text{corr}[Y_{iu}, Y_{it}] = \rho^{(ns)}_{\lvert u-t\rvert}$ $= \begin{cases} \frac{\rho\{\Sigma_{j=0}^{min(u,t)-1}(-\rho)^j\mu_{i,min(u,t)-j}\}}{\sqrt{\mu_{iu}\mu_{it}}} & \text{for } \lvert u-t\rvert = 1 \\ 0 & \text{otherwise,} \end{cases}$
EQC	$y_{it} = \rho * y_{i1} + d_{it}, t = 2,\ldots,T$ $y_{i1} \sim Poi(\mu_{i1})$ $d_{it} \sim P(\mu_{it} - \rho\mu_{i1}), t = 2,\ldots,T$	$E[Y_{it}] = \mu_{it}$ $\text{var}[Y_{it}] = \mu_{it}$ $\text{corr}[Y_{iu}, Y_{it}] = \rho^{(ns)}_{\lvert u-t\rvert}$ $= \frac{\rho\mu_{i1}}{\sqrt{\mu_{iu}\mu_{it}}}$

6.5.2 Estimation of Parameters

It follows from Sections 6.5.1.1 – 6.5.1.3 (see also Table 6.9) that all three nonstationary, namely AR(1), MA(1), and EQC, models have the same mean and variance structures. Their correlation structures are, however, different; that is, the nonstationary correlation matrix $C_i^{(ns)}(x_i,\rho) = (c_{i,ut}^{(ns)}(x_{iu},x_{it},\rho))$ is not the same under all three models. Suppose that the structure is identified (see Section 6.5.3 for an exploratory way for the model selection). Now assuming that we have a consistent estimate for ρ, say $\hat{\rho}$, we may obtain a consistent and highly efficient estimate for β by using the GQL approach that we provide below.

GQL Estimating Equation for β: Similar to the GQL estimation (6.26) for the stationary case, we now solve the GQL estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{(ns)^{-1}}(\hat{\rho})(y_i - \mu_i) = 0, \tag{6.56}$$

where $\mu_i = (\mu_{i1},\ldots,\mu_{it},\ldots,\mu_{iT})'$ is the mean vector of $y_i = (y_{i1},\ldots,y_{it},\ldots,y_{iT})'$ with

$$\begin{aligned} \mu_{it} &= \exp(x_{it}'\beta) \\ \Sigma_i^{(ns)}(\hat{\rho}) &= A_i^{1/2} C_i^{(ns)}(x_i,\hat{\rho}) A_i^{1/2}, \end{aligned} \tag{6.57}$$

where

$$A_i = \text{diag}[\sigma_{i11},\ldots,\sigma_{itt},\ldots,\sigma_{iTT}],$$

with $\sigma_{itt} = \exp(x_{it}'\beta)$. Furthermore, in (6.56), $\partial \mu_i'/\partial \beta = X_i'A_i$, with X_i as the $T \times p$ covariate matrix as defined earlier.

Let $\hat{\beta}_{GQL}$ denote the solution of (6.56) after using $\hat{\rho}$ computed under the selected model. Under mild regularity conditions one may then show that $\hat{\beta}_{GQL}$ has the asymptotic (as $K \to \infty$) normal distribution given by

$$K^{1/2}(\hat{\beta}_{GQL} - \beta) \sim N\left(0,\ K\left[\sum_{i=1}^{K} X_i' A_i \Sigma_i^{(ns)^{-1}} A_i X_i\right]^{-1}\right).$$

We now show how to compute $\hat{\rho}$ under all three models.

6.5.2.1 Estimation of ρ Parameter Under AR(1) Model

Moment Equation for ρ: Under the nonstationary AR(1) model (6.44), the moment estimate of ρ has the formula given by

$$\hat{\rho} = \frac{\sum_{i=1}^{K}\sum_{t=2}^{T} \tilde{y}_{it}\tilde{y}_{i,t-1}}{\sum_{i=1}^{K}\sum_{t=1}^{T} \tilde{y}_{it}^2} \frac{KT}{\sum_{i=1}^{K}\sum_{t=2}^{T}[\mu_{i,t-1}/\mu_{it}]^{1/2}}, \tag{6.58}$$

where $\tilde{y}_{it} = [y_{it} - \mu_{it}]/\sqrt{\mu_{it}}$. Note that the formula for ρ given by (6.58) was obtained by equating the lag 1 sample autocorrelation with its population counterpart given by (6.46). Furthermore, $\hat{\rho}$ computed by (6.58) must satisfy the range restriction given in (6.47). This implies that if the value of $\hat{\rho}$ computed by (6.58) falls beyond the range shown in (6.47), we use the upper limit of ρ given in (6.47) as the estimate of ρ.

6.5.2.2 Estimation of ρ Parameter Under MA(1) Correlation Model

Note that unlike the formula for lag 1 correlations (6.46) under the AR(1) model, the formula for this lag 1 correlation given by (6.50) under the nonstationary MA(1) model (6.49) involves a complicated summation. Thus, it is convenient to solve the moment equation for ρ by using the Newton−Raphson iterative technique. To be specific, by writing the moment equation as

$$g(\rho) = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-1}\tilde{y}_{it}\tilde{y}_{i,t+1}/K(T-1)}{\sum_{i=1}^{K}\sum_{t=1}^{T}\tilde{y}_{it}^{2}/KT} - \frac{\rho}{T-1}\sum_{u=1}^{T-1}\left[\frac{\sum_{j=0}^{u-1}(-\rho)^{j}\mu_{i,u-j}}{\sqrt{\mu_{iu}\mu_{i,u+1}}}\right] = 0, \tag{6.59}$$

we solve for ρ iteratively by using the Newton−Raphson iterative formula

$$\hat{\rho}(r+1) = \hat{\rho}(r) - \left[\{\frac{\partial g(\rho)}{\partial \rho}\}^{-1} g(\rho)\right]_{(r)},$$

where $[\cdot]_{(r)}$ denotes that the expression within brackets is evaluated at $\rho = \hat{\rho}(r)$, the rth iterative value of ρ. Note that $\hat{\rho}$ must satisfy the range restriction (6.51).

6.5.2.3 Estimation of ρ Parameter Under Exchangeable (EQC) Correlation Model

The moment estimating equation for the ρ parameter for the exchangeable model is quite similar to that of the AR(1) model. The difference between the two equations is that under the AR(1) process we have considered all lag 1 standardized residuals, whereas under the exchangeable model, one needs to use standardized residuals of all possible lags. Thus, following (6.58) for the AR(1) model, we write the moment formula for ρ under the exchangeable model as

$$\hat{\rho} = \frac{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\tilde{y}_{it}\tilde{y}_{i,t+\ell}}{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\tilde{y}_{it}^{2}} \frac{KT}{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\frac{\mu_{i1}}{[\mu_{it}\mu_{i,t+\ell}]^{\frac{1}{2}}}}, \tag{6.60}$$

where $\tilde{y}_{it} = [y_{it} - \mu_{it}]/\sqrt{\mu_{it}}$. Note that $\hat{\rho}$ must satisfy the range restriction in (6.55). This implies that if the value of $\hat{\rho}$ computed by (6.58) falls beyond the range shown in (6.55), we take $\hat{\rho}$ as the upper limit of ρ given in (6.55).

6.5.3 Model Selection

Note that in the stationary case it is not necessary to identify the correlation structure for the construction of the estimating equation (6.26) for β. This is because the estimating equation (6.26) is constructed based on a common correlation structure for $C_i^*(\rho)$ as given by (6.25) with ρ_ℓ estimated as

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\tilde{y}_{it}\tilde{y}_{i,t+\ell}/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}\tilde{y}_{it}^2/KT}, \tag{6.61}$$

(see also (6.27)) where $\tilde{y}_{it} = [y_{it}-\mu_{it}]/\sqrt{\sigma_{itt}}$. Nevertheless, if one would like to identify the stationary correlation structure for the purpose of forecasting or other reasons, this could be done by using the values of $\hat{\rho}_\ell$ for $\ell = 1,\ldots,T-1$. This is because one may show that

$$E[\hat{\rho}_\ell] = \rho_\ell,$$

approximately, and it is reasonable to use the values of $\hat{\rho}_\ell$ for $\ell = 1,\ldots,T-1$, to identify a stationary correlation structure.

As far as the identification of a nonstationary correlation structure is concerned, it appears that the values of $\hat{\rho}_\ell$ can still be used for such an identification. More specifically, simply compute the values of $\hat{\rho}_\ell$ by (6.61) and compare their pattern for best possible matching with those of $E[\hat{\rho}_\ell]$ under desired models for all possible values of $\rho = 0.0,\ 0.05,\ldots,0.90,\ 0.95$. Suppose that it is intended to find out whether the longitudinal count data follow one of the low-order, namely AR(1), MA(1), or EQC, models. To resolve such an issue, one would compute the $E[\hat{\rho}_\ell]$ under all these three models and select that model which produces a pattern for $\hat{\rho}_\ell$ similar to that of $E[\hat{\rho}_\ell]$.

For the longitudinal count data, the formulas for the expectations under the AR(1), MA(1), or EQC models are given by

$$\text{For } AR(1): E[\hat{\rho}_\ell] = \frac{\rho^\ell}{K(T-\ell)}\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\left[\frac{\mu_{it}}{\mu_{i,t+\ell}}\right]^{1/2} \quad \text{for } \ell = 1,\ldots,T-1 \tag{6.62}$$

$$\text{For } MA(1): E[\hat{\rho}_\ell] = \begin{cases} \frac{\rho}{K(T-\ell)}\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\left[\frac{\sum_{j=0}^{t-1}(-\rho)^j\mu_{i,t-j}}{\sqrt{\mu_{it}\mu_{i,t+\ell}}}\right] & \text{for } \ell = 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.63}$$

$$\text{For } EQC: E[\hat{\rho}_\ell] = \frac{\rho}{K(T-\ell)}\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\left[\frac{\mu_{i1}}{\{\mu_{it}\mu_{i,t+\ell}\}^{\frac{1}{2}}}\right], \tag{6.64}$$

for $\ell = 1,\ldots,T-1$, where $\mu_{it} = \exp(x'_{it}\beta)$ for all $t = 1,\ldots,T$. Note that as far as the value of β is concerned for computing $\hat{\rho}_\ell$ by (6.61) and the expectations by (6.62) – (6.64), this may be obtained by solving the GQL estimating equation (6.26) under the 'working' independence assumption $\rho = 0.0$. This is because such an estimate is always consistent and one does not necessarily require an efficient estimate for β before the correlation structure is identified.

Further note that if the time dependent covariates are not so different over time, then the expected values in (6.62) – (6.64) would almost agree with the correlation pattern under the stationary case, described through (6.17), (6.19), and (6.22). To demonstrate this, we now examine empirically the pattern for $E[\hat{\rho}_\ell]$ under all three correlation models. For this purpose, we consider two time-dependent covariates as follows:

$$x_{it1} = \begin{cases} \frac{1}{2} & \text{for } t = 1,\ 2; \quad i = 1,\ldots,K/4 \\ 1 & \text{for } t = 3,\ 4; \quad i = 1,\ldots,K/4 \\ -\frac{1}{2} & \text{for } t = 1; \quad i = (K/4)+1,\ldots,3K/4 \\ 0 & \text{for } t = 2,\ 3; \quad i = (K/4)+1,\ldots,3K/4 \\ \frac{1}{2} & \text{for } t = 4; \quad i = (K/4)+1,\ldots,3K/4 \\ \frac{t}{8} & \text{for } t = 1,\ldots,4; \ i = (3K/4)+1,\ldots,K, \end{cases}$$

and

$$x_{it2} = \begin{cases} \frac{t-2.5}{8} & \text{for } t = 1,\ldots,4; \ i = 1,\ldots,K/2 \\ 0 & \text{for } t = 1,2; \quad i = (K/2)+1,\ldots,K \\ \frac{1}{2} & \text{for } t = 3,\ 4; \quad i = (K/2)+1,\ldots,K. \end{cases}$$

For $T = 4$ and $K = 100$, the values for $E[\hat{\rho}_\ell]$ computed by (6.62) – (6.64) for suitable values of ρ are displayed in Table 6.10.

It is clear from the results of the table that the $E[\hat{\rho}_\ell]$ for $\ell = 1,\ldots,T-1$, exhibit an exponentially decaying pattern under the nonstationary AR(1) model, whereas they exhibit a truncated pattern under the MA(1) model, and a constant pattern under the EQC model. These patterns are quite similar to those under the respective stationary correlation structure. Thus, it appears that in practice one may still exploit the values of $\hat{\rho}_\ell$ computed by (6.61) in order to diagnose the nonstationary correlation pattern. More specifically, because the values of $E[\hat{\rho}_\ell]$ for $\ell = 1,\ldots,T-1$, under the AR(1), MA(1), and EQC models exhibit three different patterns, and because the values of $\hat{\rho}_\ell$ computed from the data should reflect the pattern supported by the values of $E[\hat{\rho}_\ell]$, it is quite reasonable to examine the pattern generated by the values of $\hat{\rho}_\ell$ to diagnose the appropriate model.

Table 6.10 The pattern for $E[\hat{\rho}_\ell]$ for lag $\ell = 1, \ldots, T-1$, under AR(1), MA(1), and EQC correlation structures for longitudinal count data with selected values for the correlation index parameter ρ.

Correlation Structure								
AR(1)			MA(1)			EQC		
ρ	ℓ	$E[\hat{\rho}_\ell]$	ρ	ℓ	$E[\hat{\rho}_\ell]$	ρ	ℓ	$E[\hat{\rho}_\ell]$
0.3	1	0.282	0.1	1	0.089	0.3	1	0.251
	2	0.078		2	0.0		2	0.248
	3	0.022		3	0.0		3	0.248
0.5	1	0.469	0.2	1	0.168	0.5	1	0.417
	2	0.216		2	0.0		2	0.413
	3	0.103		3	0.0		3	0.412
0.6	1	0.563	0.3	1	0.239	0.6	1	0.502
	2	0.312		2	0.0		2	0.495
	3	0.178		3	0.0		3	0.494
0.68	1	0.638	0.4	1	0.306	0.7	1	0.587
	2	0.400		2	0.0		2	0.577
	3	0.259		3	0.0		3	0.577

6.6 More Nonstationary Correlation Models

6.6.1 Models with Variable Marginal Means and Variances

In this section, we demonstrate that as opposed to the nonstationary MA(1) model in (6.49), one may construct a different MA(1) model that produces the mean and the variance functions different from those produced by the nonstationary AR(1) (6.44) and EQC (6.53) models. These two latter models in (6.44) and (6.53) produce the mean and the variance as

$$E[Y_{it}] = \text{var}[Y_{it}] = \exp(x'_{it}\beta). \tag{6.65}$$

We now construct an alternative MA(1) model to (6.49), and examine its mean, variance, and correlation structures.

6.6.1.1 Nonstationary MA(1) Models

Suppose that the non-stationary MA(1) model for the count responses has the same form, that is,

$$y_{it} = \rho * d_{i,t-1} + d_{it}, \tag{6.66}$$

as in (6.18) under the stationary case, but the model components are now assumed to satisfy the following distributional assumptions.

Assumption 1. For $t = 1, \ldots, T$, the discrete errors d_{it} follow the Poisson distribution as $d_{it} \sim P(\mu_{it}/(1+\rho))$, with $\mu_{it} = \exp(x'_{it}\beta)$.

Assumption 2. For all $t = 1, \ldots, T$, d_{it}s are independent.

Assumption 3. An initial discrete error $d_{i0} \sim P(\mu_{i0}/[1+\rho])$, where the choice of μ_{i0}, a function of some initial or past covariates, is left to the user. In the stationary case, $\mu_{i0} = \mu_{i1} = \cdots = \mu_{iT} = \mu_{i\cdot}$.

For $t = 1, \ldots, T$, by writing $z_{i,t-1} = \rho * d_{i,t-1}$, for convenience, one may now use the model (6.66) and compute the mean $\nu_{it} = E(Y_{it})$ and the variance $\sigma_{itt} = \text{var}(Y_{it})$ as

$$\nu_{it} = E_{d_{i,t-1}} E[z_{i,t-1}] + E[d_{it}] = [\rho\mu_{i,t-1} + \mu_{it}]/(1+\rho), \tag{6.67}$$

and

$$\begin{aligned}
\sigma_{itt} &= \text{var}_{d_{i,t-1}} E[z_{it}|d_{i,t-1}] + E_{d_{i,t-1}} \text{var}[z_{it}|d_{i,t-1}] + \text{var}[d_{it}] \\
&= \text{var}_{d_{i,t-1}}[\rho d_{i,t-1}] + E_{d_{i,t-1}}[\rho(1-\rho)d_{i,t-1}] + [\mu_{it}/(1+\rho)] \\
&= [\rho\mu_{i,t-1} + \mu_{it}]/(1+\rho),
\end{aligned} \tag{6.68}$$

respectively. Thus, it is clear that for $t = 1, \ldots, T$, y_{it} has the mean ν_{it} and the variance $\sigma_{itt} = \nu_{it}$, which are, however, different from the mean and the variance functions given in (6.65) under the AR(1) and EQC models. Also, it is to be noted that the ρ parameter in the MA(1) model (6.66) must satisfy the range restriction $\max[-\mu_{it}/\mu_{i,t-1}] < \rho < 1$, for all i and t. Next by similar calculations as in the AR(1) model, it follows from (6.67) – (6.68) that under the MA(1) model, the ℓth $\ell = 1, \ldots, T-1$, lag autocorrelation is given by

$$\text{corr}(Y_{it}, Y_{i,t-\ell}) = c^{(ns)}_{it,t-l}(x_i, \rho) = \left\{ \begin{array}{ll} [\rho\mu_{i,t-\ell}/(1+\rho)]/[\nu_{it}\nu_{i,t-\ell}]^{1/2} & \text{for } \ell = 1 \\ 0 & \text{for } \ell > 1. \end{array} \right\}, \tag{6.69}$$

which is nonstationary. This correlation structure is different from that (6.50) of the other MA(1) model (6.49).

Thus, under this alternative nonstationary MA(1) model (6.66), it is not only that the correlations are different from those of the AR(1) and EQC models, but the mean and the variances are also different.

6.6.2 Estimation of Parameters

Note that the three nonstationary models, namely AR(1), MA(1), and EQC introduced in Sections 6.5.1.1, 6.5.1.2, and 6.5.1.3, respectively, produce the same mean and variance functions but different correlation structures. In spite of their different correlation structures, the regression parameter β was estimated by solving the GQL estimating equation (6.56), which is unbiased for zero vector, irrespective of the model for the data. This happens because all three correlation models produce the same mean vector μ_i as given in (6.56). As opposed to Section 6.5, in Section 6.6 we now assume that the repeated count data are generated following either the AR(1) (6.44) or EQC (6.53) model from Section 6.5, or following the MA(1) model (6.66) introduced in Section 6.6.1.1. The MA(1) model (6.66) produces different mean and variance structure, thus it is no longer possible to use the estimating equation (6.56) for β to obtain consistent estimate, under the MA(1) model (6.66). This is, however, a valid equation to solve for β under the AR(1) and EQC models. Furthermore, for these two models (6.44) and (6.53), the ρ parameter is consistently estimated by (6.58) and (6.60), respectively.

In the next section, we demonstrate how to estimate β and ρ parameters of the MA(1) model (6.66).

6.6.2.1 GQL Estimation for Regression Effects β

We now fit the nonstationary MA(1) model (6.66) to the longitudinal count data. The mean and the variance structures under this model are given in (6.67) – (6.68), whereas the nonstationary correlation structure is given by (6.69).

Let

$$v_i = (v_{i1}, \ldots, v_{it}, \ldots, v_{iT})'$$

be the mean vector of y_i, where for $t = 1, \ldots, T$,

$$v_{it} = [\mu_{it} + \rho \mu_{i,t-1}]/(1+\rho)$$

by (6.67). For convenience, we assume that $\mu_{i0} = 0$. Furthermore, let $\Sigma_i^{(ns)}(\rho) = (\sigma_{iut})$ be the $T \times T$ covariance matrix of y_i, where

$$\sigma_{iut} = \begin{cases} \sigma_{itt}, & \text{if } u = t \\ \frac{\rho \mu_{iu}}{1+\rho}, & \text{if } u < t, \end{cases} \tag{6.70}$$

with σ_{itt} as in (6.68). It then follows that for known ρ, one may write the GQL estimating equation for β as

$$\sum_{i=1}^{K} \frac{\partial v_i'}{\partial \beta} \Sigma_i^{(ns)^{-1}}(\hat{\rho})(y_i - v_i) = 0, \tag{6.71}$$

which is a different estimating equation from that of under the AR(1) model (6.44) and EQC model (6.53).One may now solve (6.71) iteratively by using the Newton−Raphson algorithm. To be specific, (6.71) is solved for β iteratively by using

$$\hat{\beta}(r+1) = \hat{\beta}(r) + \left[\left\{ \sum_{i=1}^{K} [(X_i' A_i + Z_i' B_i) \Sigma_i^{-1} (A_i X_i + B_i Z_i)] \right\}^{-1} \right.$$

$$\left. \times \sum_{i=1}^{K} \left\{ (X_i' A_i + Z_i' B_i) \Sigma_i^{-1} (y_i - \nu_i) \right\} \right]_{[r]}, \tag{6.72}$$

where

$$X_i' = (x_{i1}, \ldots, x_{it}, \ldots, x_{iT}), \quad Z_i' = (1_p, x_{i1}, \ldots, x_{i,T-1}),$$

$$A_i = \text{diag}(\frac{\mu_{i1}}{1+\rho}, \frac{\mu_{i2}}{1+\rho}, \ldots, \frac{\mu_{it}}{1+\rho}, \ldots, \frac{\mu_{iT}}{1+\rho}),$$

$$B = \text{diag}(0, \frac{\rho\mu_{i1}}{1+\rho}, \frac{\rho\mu_{i2}}{1+\rho}, \ldots, \frac{\rho\mu_{it}}{1+\rho}, \ldots, \frac{\rho\mu_{i,T-1}}{1+\rho}),$$

and $[.]_r$ denotes the fact that the expression within the brackets is evaluated at $\hat{\beta}(r)$. Let $\hat{\beta}_{GQL}$ denote the solution obtained from (6.72). Under mild regularity conditions it may be shown that $\hat{\beta}_{GQL}$ has the asymptotic (as $K \to \infty$) normal distribution given as

$$K^{\frac{1}{2}}(\hat{\beta}_{GQL} - \beta) \sim N\left(0,\ K \left[\sum_{i=1}^{K} (X_i' A_i + Z_i' B_i) \Sigma_i^{-1} (A_i X_i + B_i Z_i) \right]^{-1} \right). \tag{6.73}$$

6.6.2.2 Moment Estimation for the Correlation Parameter ρ

As far as the ρ parameter is concerned, we estimate this parameter consistently by using the well-known method of moments. For the purpose, we first observe under the MA(1) model that

$$E\left[\frac{(Y_{it} - \nu_{it})}{\sqrt{\nu_{it}}} \right]^2 = 1$$

$$E\left[\frac{(Y_{it} - \nu_{it})}{\sqrt{\nu_{it}}} \frac{(Y_{i,t-1} - \nu_{i,t-1})}{\sqrt{\nu_{i,t-1}}} \right] = \frac{\rho}{1+\rho} \frac{\mu_{i,t-1}}{\sqrt{\nu_{it}\nu_{i,t-1}}}. \tag{6.74}$$

Consequently, one may obtain a consistent estimator of ρ by solving the moment equation

$$\frac{a(\rho)}{b(\rho)} = \frac{\rho}{1+\rho}c(\rho), \tag{6.75}$$

where

$$a(\rho) = \frac{1}{K(T-1)}\sum_{i=1}^{K}\sum_{t=2}^{T}\frac{(Y_{it}-\nu_{it})}{\sqrt{\nu_{it}}}\frac{(Y_{i,t-1}-\nu_{i,t-1})}{\sqrt{\nu_{i,t-1}}}$$

$$b(\rho) = \frac{1}{KT}\sum_{i=1}^{K}\sum_{t=1}^{T}\left[\frac{(Y_{it}-\nu_{it})}{\sqrt{\nu_{it}}}\right]^2,$$

and

$$c(\rho) = \frac{1}{K(T-1)}\sum_{i=1}^{K}\sum_{t=2}^{T}\frac{\mu_{i,t-1}}{\sqrt{\nu_{it}\nu_{i,t-1}}}. \tag{6.76}$$

Note that unlike solving for ρ by (6.58) under the AR(1) process or by (6.60) under the EQC model, solving (6.75) for ρ under the MA(1) model is complicated as ν_{it} contains ρ for all $t = 1,\ldots,T$. One may, however, obtain an approximate solution, based on an iterative technique by using an initial value of ρ, say ρ_0, in all ν_{it}, and solving (6.75) for ρ as

$$\rho_1 = \frac{a(\rho_0)}{b(\rho_0)c(\rho_0)-a(\rho_0)}. \tag{6.77}$$

Next one may improve the estimate of ρ by using ρ_1 in place of ρ_0 in (6.75). That is, the new solution of ρ is obtained as

$$\rho_2 = \frac{a(\rho_1)}{b(\rho_1)c(\rho_1)-a(\rho_1)}. \tag{6.78}$$

This iteration continues until convergence.

6.6.3 Model Selection

Under the assumption that the longitudinal count data follow either the nonstationary AR(1) (6.44) or EQC (6.53) model described in Section 6.5, we have estimated their common regression parameter by (6.56), and their correlation parameter ρ was estimated by (6.58) and (6.60), respectively. Next, for the estimation of the parameters of the MA(1) model (6.66), we have used the GQL approach (6.71) for β estimation, and the moment estimating equation (6.75) for the estimation of the ρ parameter. Now the question arises, which model to recommend for use in practice? We consider a lag 1 model fitting approach to answer this question. Note that this model selection approach is different from that we have used in Section 6.5.3. One of the reasons for this difference in model selection approaches is that in Section 6.5 we have considered models with the same mean functions, whereas in this section we have considered models with different mean functions. To be more specific, when the models do not agree for the mean functions, it is better to fit them to the data separately and then see which model fits the data best. Thus, in this section, we fit a model M (say) to the data and simply compute the error sum of squares, G_M, under the model M, defined by

$$G_M = \sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it} - \hat{y}_{it}(M)]^2, \tag{6.79}$$

and recommend that model with the smallest value of the error sum of squares. In (6.79), $\hat{y}_{it}(M)$ denotes the fitted value of y_{it} under the model M.

The formula for $\hat{y}_{it}(M)$under each of the three models are as follows.

When Nonstationary AR(1) Model (6.44) Is Fitted

$$\hat{y}_{it} = \begin{cases} \hat{\mu}_{it} & \text{for } t = 1 \\ \hat{\mu}_{it} + \hat{\rho}\{y_{i,t-1} - \hat{\mu}_{i,t-1}\} & \text{for } t = 2,\ldots,T, \end{cases} \tag{6.80}$$

with $\hat{\mu}_{it} = \exp(x'_{it}\hat{\beta})$, where $\hat{\beta}$ is obtained by solving the GQL estimating equation (6.56) and $\hat{\rho}$ is obtained as the moment estimate by using (6.58).

When Non-stationary MA(1) Model (6.66) is Fitted

$$\hat{y}_{it} = \begin{cases} \frac{\hat{\mu}_{it}}{1+\hat{\rho}} & \text{for } t = 1 \\ \frac{\hat{\mu}_{it}+\hat{\rho}\hat{\mu}_{i,t-1}}{1+\hat{\rho}} & \text{for } t = 2,\ldots,T, \end{cases} \tag{6.81}$$

with $\hat{\mu}_{it} = \exp(x'_{it}\hat{\beta})$, but $\hat{\beta}$ is obtained by solving the GQL estimating equation (6.71) and $\hat{\rho}$ is obtained as the moment estimate by solving (6.75). Note that estimating equations in (6.71) and (6.75) under the MA(1) model are similar to but different from the AR(1) based estimating equations (6.56) and (6.58), respectively.

When Nonstationary Exchangeable or Equicorrelation (EQC) Model (6.53) Is Fitted

$$\hat{y}_{it} = \begin{cases} \hat{\mu}_{it} & \text{for } t = 1 \\ \hat{\mu}_{it} + \hat{\rho}\{y_{i1} - \hat{\mu}_{i1}\} & \text{for } t = 2,\ldots,T, \end{cases} \tag{6.82}$$

with $\hat{\mu}_{it} = \exp(x'_{it}\hat{\beta})$, where $\hat{\beta}$ and $\hat{\rho}$ are obtained by solving the GQL (6.56) and moment estimating equation (6.60).

6.6.4 Estimation and Model Selection: A Simulation Example

We now consider a simulation study and examine the performance of the GQL estimation approach discussed in Section 6.6.2. We also examine the performance of the mean squared errors (MSEs) based model selection approach discussed in Section 6.6.3. We demonstrate here that if a misspecified model is used, then the GQL approach may lead to inconsistent estimates for the regression effects causing a serious inference problem. This happens when the mean and the variance functions of the true model are different from those of the so-called ‘working’ or misspecified model.

6.6.4.1 Simulated Estimates Under the True and Misspecified Models

To choose a simulation design, we take $p = 2$ and $\beta_1 = \beta_2 = 0.5$. With regard to the correlation index parameter, we consider two cases, one with moderately large $\rho = 0.5$ and the other with large $\rho = 0.75$. Next we choose $K = 300$, where K is the number of independent individuals. As far as the values of the covariates are concerned, we consider two time-dependent covariates given in Section 6.5.3.

Next, for a selected value of K, and ρ, we simulate the longitudinal responses $y_{i1}, \ldots, y_{iT}$, following a true, say AR(1) or exchangeable correlation model as described in Section 6.5.1, or the MA(1) model as described in Section 6.6.1. We consider 1000 simulations. In each simulation, we then estimate the parameters β_1, β_2, and ρ, by using the formulas for all three processes as discussed in Section 6.6.2. The simulated mean and the simulated standard error of the estimates are reported in Table 6.11.

The results in Table 6.11 clearly indicate that fitting a ‘working’ nonstationary model can be extremely dangerous. For example, when the longitudinal data are generated, say following the MA(1) model, and also the estimates are obtained by fitting the MA(1) model, the GQL estimates appear to perform very well. The GQL estimates computed based on either the AR(1) or EQC model, however, appear to be far off from the true parameter values. To be specific, when $\rho = 0.75$, the true MA(1) based GQL estimates for $\beta_1 = 0.5$ and $\beta_2 = 0.5$ are 0.491 with standard error 0.175, and 0.499 with standard error 0.175, respectively. These estimates are very close to the true values. Similarly, the moment estimate for $\rho = 0.75$ is found to be 0.749 with small standard error 0.064, which indicates superb performance of the GQL approach provided the true model is used for the estimation. On the contrary, when AR(1) model is used as the ’working’ model, the regression estimates are found to be -1.016 and 1.709 for true $\beta_1 = \beta_2 = 0.5$. It is clear that these estimates are complete nonsense. Similar results hold for ρ estimation. The AR(1) based moment estimate for $\rho = 0.75$ is found to be 1.000, which is also highly biased. Note that these results are not surprising. This is because unlike under the stationary models [Liang and Zeger (1986), Sutradhar (2003)], the mean and variance structures under different correlation models may be different.

Table 6.11 The simulated means and the simulated standard errors of the estimates of the regression and the correlation index parameters under both true and 'working' nonstationary AR(1), MA(1), and EQC (equicorrelations) models for longitudinal count data, with true $\beta_1 = \beta_2 = 0.5$, for $K = 300$ individuals, and a selected value of ρ, based on 1000 simulations.

			True Nonstationary Correlation Model					
			AR(1)		MA(1)		EQC	
Working Model	True ρ	Parameters	SM	SSE	SM	SSE	SM	SSE
AR(1)	0.60	β_1	0.499	0.111	−0.159	0.370	0.502	0.125
		β_2	0.494	0.103	1.171	0.306	0.491	0.116
		ρ	0.599	0.033	0.847	0.076	0.504	0.044
	0.75	β_1	0.499	0.094	−1.016	0.279	0.504	0.114
		β_2	0.503	0.087	1.790	0.232	0.499	0.104
		ρ	0.749	0.029	1.000	0.004	0.696	0.042
MA(1)	0.60	β_1	0.477	0.138	0.483	0.178	0.360	0.130
		β_2	0.388	0.133	0.506	0.177	0.601	0.129
		ρ	0.386	0.031	0.598	0.062	0.249	0.039
	0.75	β_1	0.481	0.127	0.491	0.175	0.368	0.122
		β_2	0.367	0.125	0.499	0.175	0.611	0.121
		ρ	0.452	0.028	0.749	0.064	0.291	0.042
EQC	0.60	β_1	0.498	0.126	0.215	0.278	0.498	0.110
		β_2	0.496	0.111	0.875	0.253	0.498	0.097
		ρ	0.521	0.042	0.717	0.080	0.597	0.044
	0.75	β_1	0.497	0.115	0.777	0.446	0.498	0.090
		β_2	0.500	0.097	1.618	0.350	0.500	0.080
		ρ	0.655	0.038	0.966	0.054	0.749	0.041

Remark that because the AR(1) and EQC models produce the same mean and the variance functions, the estimates under model misspecification do not vary too much but the standard errors tend to be larger under the misspecified models [Sutradhar and Das (1999)]. For example, when the data are generated following the AR(1) model, the AR(1) model based estimates for β_1, β_2, and ρ, have the standard errors 0.094, 0.087, 0.029, whereas the EQC model based corresponding standard errors are 0.115, 0.097, 0.038, confirming inefficient estimation under the 'working' correlation models.

In summary, when the longitudinal data follow a nonstationary correlation model, the effect of selecting a 'working' model with different mean and variance functions can be very serious. Thus, it is important to identify the true model to fit the data.

6.6.4.2 Model Selection

Note that it is practical to attempt to fit a possible low-order correlation model to given longitudinal data. But it may not be easy to identify the actual correlation structure for the data, especially when the data may follow one of the three nonstationary correlation models discussed in the paper. We thus recommend fitting all three models initially to the given data and compute the G_M statistic defined in

(6.79) under all three fitted models. One may then choose the model which produces the smallest value of the statistic G_M. The simulation results reported in Table 6.12 appear to support this technique of model selection.

Table 6.12 The simulated error sum of squares (ESS) under both true and 'working' nonstationary AR(1), MA(1), and EQC (equi-correlations) models for longitudinal count data, with true $\beta_1 = \beta_2 = 0.5$, for $K = 300$ individuals, and a selected value of ρ, based on 1000 simulations.

	True nonstationary Correlation Model			
		AR(1)	MA(1)	EQC
Selected ρ	Working Model	ESS	ESS	ESS
0.60	AR(1)	0.967	1.378	1.180
	MA(1)	1.281	1.138	1.158
	EQC	1.053	1.347	1.012
0.75	AR(1)	0.788	1.450	1.046
	MA(1)	1.249	1.120	1.145
	EQC	0.919	1.425	0.856

For example, when the data were generated following the nonstationary AR(1) model (6.44) with $\rho = 0.75$, the simulated average values of the G_M statistic computed by using the fitted values based on AR(1) (6.80), MA(1) (6.81), and EQC (6.82) models are found to be 0.788, 1.450, and 1.046, respectively. It is then clear that when the data follow the AR(1) model and the AR(1) model is fitted, the G_M statistic has the smallest value. Similar results hold under the other two models too.

6.7 A Data Example: Analyzing Health Care Utilization Count Data

We now consider an illustration for the application of the nonstationary correlation models for repeated count data discussed in Section 6.6, by analyzing the health care utilization data, earlier studied by Sutradhar (2003), for example. This dataset, provided in Appendix 6A, is a part of the longitudinal dataset collected by the General Hospital of the city of St. John's, Canada. To be specific, here we consider the longitudinal count data that contain the complete records for 144 individuals for four years $(n = 4)$ from $1985 - 1988$. The number of visits to a physician by each individual during a given year was recorded as the response, and this was repeated for four years. Also, the information on four covariates, namely, gender, number of chronic conditions, education level, and age, were recorded for each individual. Note that as the responses are counts, it is appropriate to assume that the response variable, marginally, follows the Poisson distribution, and the repeated counts recorded for four years will be longitudinally correlated. Along the lines of

Liang and Zeger (1986) we assume that the data may follow any of the low-order correlations such as AR(1), MA(1), or EQC models discussed in Section 6.6. Note that because these models produce different mean and the variance structures, they must be fitted by using these varied mean, variance, and correlation structures for the purpose of obtaining consistent and efficient estimates for the regression effects β and the correlation index parameter ρ.

Following the notations used in Sections 6.5 and 6.6, the four covariates for the ith $(i = 1,\ldots,K = 144)$ individual at time t $(t = 1,\ldots,4)$ are denoted by $x_{it1}, x_{it2}, x_{it3},$ and x_{it4} respectively. The first covariate geneder was coded as 0 for female and 1 for male. Thus, at any time t, $x_{it1} = 0$ if the ith individual is female, otherwise $x_{it1} = 1$. Similarly, the number of chronic diseases was coded as $x_{it2} = 0$ for the absence of chronic disease for the ith individual at time t, and $x_{it2} = 1$ if the ith individual had 1 or more chronic diseases at time t. The third covariate, education level, x_{it3}, was coded as 1 for less than high school, and 0 for high school or higher education. The last covariate, x_{it4}, represents the age of the individual. The effects of these covariates are denoted by $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$, so that the mean of the count response for the ith individual at a time point t is given by (6.65) under the nonstationary AR(1) and EQC structures, and by (6.67) under the nonstationary MA(1) model. In all these mean functions $x_{it} = (x_{it1}, x_{it2}, x_{it3}, x_{it4})'$.

We now apply the GQL estimation methodology discussed in Section 6.6. By using the response data y_{it} and x_{it} vector for all $i = 1,\ldots,144$, individuals and over $t = 1,\ldots,4$, years, we obtain the estimate of β and ρ from Section 6.5.1.1 under the nonstationary AR(1), from Section 6.6.1.1 under the MA(1), and similarly from Section 6.5.1.3 under the EQC models. These results along with the standard errors of the estimates of β computed by using the asymptotic covariance matrices from these three sections, are reported in Table 6.13.

Table 6.13 Comparison of the estimates of the regression and the correlation parameters under the nonstationary AR(1), MA(1), and EQC (equicorrelations) models in fitting the health care utilization data.

Nonstationary Correlation Models

	AR(1)		MA(1)		EQC	
Parameters	EST	SE	EST	SE	EST	SE
Gender effect (β_1)	−0.223	0.060	−0.179	0.054	−0.204	0.065
Chronic effect (β_2)	0.374	0.072	0.363	0.065	0.341	0.078
Education effect (β_3)	−0.428	0.074	−0.400	0.066	−0.390	0.081
Age effect(β_4)	0.029	0.001	0.031	0.001	0.029	0.001
ρ	0.554	–	0.769	–	0.529	–
$\rho_y(1)$	0.546	–	0.486	–	0.521	–
G_M	14.20	–	20.46	–	15.34	

As far as the selection of a model from these three lower-order models is concerned, we have computed the fitted residual squared distance G_M by (6.79) under all three models and reported them in the same Table 6.13. As the G_M statistic has the lowest value 14.20 under the AR(1) structure, we chose the AR(1) model to interpret the estimates.

As the first covariate gender was coded as 1 for male and 0 for female, it follows from (6.65) and (6.67) that the negative value of $\hat{\beta}_1 = -0.223$ suggests that the females made more visits to the physician as compared to the males. The positive values of $\hat{\beta}_2 = 0.374$ and $\hat{\beta}_4 = 0.029$ suggest that individuals having one or more chronic diseases or individuals belonging to the older age group pay more visits to the physicians, as expected. The third covariate education level was coded as 1 for less than high school, 0 for higher education. The effect of the education level on the physician visits was found to be $\hat{\beta}_3 = -0.428$. This negative estimate shows that highly educated individuals pay more visits as compared to individuals with a low level of education. One of the reasons for this type of behavior of this covariate may be that the individuals with a high-level education (more than high school) are more concerned about their health condition as compared to the individuals with low-level education.

Note that the standard errors of the regression estimates under the AR(1) model were found to be

$$\text{s.e.}(\hat{\beta}_1) = 0.060, \quad \text{s.e.}(\hat{\beta}_2) = 0.072, \quad \text{s.e.}(\hat{\beta}_3) = 0.074, \quad s.e.(\hat{\beta}_4) = 0.001.$$

As these standard errors are quite small as compared to the corresponding values of the regression estimates, all four covariates appear to have significant effects on the physician visits. Further note that the standard errors of the estimates under the MA(1) model appear to be smaller than the corresponding standard errors under the AR(1) model. Nevertheless, the estimates under the MA(1) model cannot be trusted as it is evident from the simulation study (see Table 6.11) that they can be highly biased when the data really follow the AR(1) model. Here the data as mentioned earlier appear to follow the AR(1) model with the smallest G_M value.

6.8 Models for Count Data from Longitudinal Adaptive Clinical Trials

In a clinical trial study with human subjects, it is highly desirable that one use certain data-dependent treatment allocation rules which exploit accumulating past information to assign individuals to treatments so that more study subjects are assigned to the better treatment. For example, consider a clinical trial study to examine the performance of a new treatment for asthma prevention. Suppose that one individual patient is assigned to one of the treatments in an adaptive way and number of asthma attacks for a week is recorded. Here the number of asthma attacks for a week may be considered to follow a Poisson distribution. Once the outcome of the first individual is known, the treatment for the second individual may be decided based on the outcome of the first individual as well as the covariate information of the individual.

Similarly, a treatment is assigned to the third individual based on the outcomes of the past two individuals and their covariate information. This adaptive procedure continues for a large number of weeks, say for 100 weeks for the treatment of 100 individuals. Note that 100 or more weeks is a reasonable duration for the completion of an intensive clinical trial study. Here, the purpose is to determine the effects of the treatments after treating a large proportion of subjects by the better treatment.

Note that there are many clinical studies including the aforementioned asthma study where it may be necessary to record the count responses repeatedly over a small period of time, from a patient based on the same assigned treatment, assignment of treatment being done in a longitudinal adaptive way. For example, for the asthma problem, it may be better to collect responses from a patient weekly for a period of $T = 4$ weeks, say, where the responses will be longitudinally correlated. As far as the treatment assignment is concerned, the assignment of the treatment to the third patient, for example, will be benefitted from the first week's response of the second patient, and the first and second weeks' responses from the first patient, and so on. The main purpose of this section is to discuss such longitudinal count data collected from a clinical trial study based on a suitable adaptive design. For the purpose, following Sutradhar and Jowaheer (2006), we first provide two longitudinal adaptive designs in Section 6.8.1. In Section 6.8.2, we demonstrate through a simulation study that the longitudinal adaptive designs discussed in Section 6.8.1 indeed allocate more patients to a better treatment. The overall treatment effects and the effects of other possible covariates are consistently and efficiently estimated in Section 6.8.3 by using a weighted GQL (WGQL) approach, based on the complete data collected from all patients during the study. We remark here that the WGQL approach indicates that the longitudinal adaptive design weights responsible for the collection of the longitudinal count data are incorporated in the so-called GQL approach discussed in the previous sections.

6.8.1 Adaptive Longitudinal Designs

Autocorrelated Poisson Model Conditional on Design Weights: Suppose that K independent patients will be treated in the clinical study and T longitudinal count responses will be collected from each of them. Also, for simplicity, let there be two treatments A and B to treat these patients and A is the better treatment between the two. Next suppose that δ_i refers to the selection of the treatment for the ith $(i = 1, \ldots, K)$ patient, and

$$\delta_i = \begin{cases} 1, & \text{if } i\text{th patient is assigned to } A \\ 0, & \text{if } i\text{th patient is assigned to } B \end{cases}$$

with

$$\Pr(\delta_i = 1) = w_i \ \text{ and } \ \Pr(\delta_i = 0) = 1 - w_i. \tag{6.83}$$

Here w_i refers to the better treatment selection probability for the ith patient. Now to construct a longitudinal adaptive design one needs to derive the formulas for the selection probabilities $w_i(i=1,\ldots,K)$ so that in the long run more patients are treated by A.

Note that the value of δ_i determines the treatment by which the ith patient will be treated. Now suppose that conditional on δ_i, y_{it} denotes the count response recorded from the ith patient at time $t(t=1,\ldots,T)$, and x_{it} denotes the p-dimensional covariate vector corresponding to y_{it}, defined as

$$\begin{aligned} x_{it} &= (\delta_i, x_{it2}, \ldots, x_{itu}, \ldots, x_{itp})' \\ &= (\delta_i, x^{*'}_{it})', \end{aligned} \tag{6.84}$$

where $x^*_{it} = (x_{it2}, \ldots, x_{itu}, \ldots, x_{itp})'$ denote the $p-1\times 1$ vector of covariates such as prognostic factors (e.g., age, chronic conditions, and smoking habit) for the ith patient available at time point t. Thus, for $i=2,\ldots,K$, the distribution of δ_i, that is, the formula of w_i, will depend on $\{\delta_1,\ldots,\delta_{i-1}\}$ and available responses y_{kv} $(k=1,\ldots,i-1; 1\leq v\leq T)$ along with their corresponding covariate vector x_{kv}. For $i=1$, w_1 is assumed to be known.

As far as the availability of the repeated responses is concerned, we assume that for all $i=1,\ldots,K$, once δ_i becomes known, the repeated count responses from the ith patient will be available following a Poisson distribution with conditional mean and variance (conditional on δ_i) given by

$$E(Y_{it}|\delta_i, x^*_{it}) = \text{var}(Y_{it}|\delta_i, x^*_{it}) = \exp(\theta_{it}), \tag{6.85}$$

where $\theta_{it} = x'_{it}\beta$, with $x_{it} = (\delta_i, x^{*'}_{it})'$. Also we assume that the pairwise longitudinal correlations between two repeated count responses are given by

$$\begin{aligned} \text{corr}[(Y_{it}, Y_{iv})|\delta_i, x^*_{it}, x^*_{iv}] &= \rho^{(ns)}_{|t-v|}(\delta_i, x^*_{it}, x^*_{iv}, \rho) \\ &= c^{(ns)}_{i,tv}(\delta_i, x^*_{it}, x^*_{iv}, \rho), \end{aligned} \tag{6.86}$$

where $c^{(ns)}_{i,tv}(\delta_i, x^*_{it}, x^*_{iv}, \rho)$ has the formulas given by (6.46), (6.50), and (6.55) under the nonstationary AR(1), MA(1), and EQC models, respectively. It then follows by (6.85) and (6.86) that the conditional (on δ_i) covariance between y_{it} and y_{iv} is given by

$$\text{cov}[(Y_{it}, Y_{iv})|\delta_i, x^*_{it}, x^*_{iv}] = \rho^{(ns)}_{|t-v|}\{\exp(\theta_{it}+\theta_{iv})\}^{\frac{1}{2}}.$$

Note, however, that for simplicity we use the stationary correlations based covariance matrix given by

$$\begin{aligned} \text{cov}[(Y_{it}, Y_{iv})|\delta_i, x^*_{it}, x^*_{iv}] &\simeq c^*_{i,tv}(\rho)\{\exp(\theta_{it}+\theta_{iv})\}^{1/2} \\ &= \rho_{|t-v|}\{\exp(\theta_{it}+\theta_{iv})\}^{\frac{1}{2}}. \end{aligned} \tag{6.87}$$

6.8.1.1 Simple Longitudinal Play-the-Winner (SLPW) Rule to Formulate w_i

Note that in the cross-sectional setup, i.e., when $T = 1$ there exist a number of options to formulate the adaptive design weights w_i for $i = 1, \ldots, K$. For example, we refer to the

(i) randomized play the winner (RPW) rule [Zelen (1969); Wei and Durham (1978); Wei et al. (1990)],
(ii) random walks rule [Durham and Flournoy (1994)],
(iii) group sequential test [Jennison and Turnbull (2001)], and
(iv) optimum biased coin designs [Pocock and Simon(1975); Smith (1984); Atkinson (1999)].

The purpose of these designs is to assign a better treatment to an incoming patient based on the past outcomes of the experiment as well as the covariate information. Note that even if there are controversies [Royall 1991; Farewell, Viveros, and Sprott (1993)] about the usefulness of the play the winner rule, this seems to be the only design which was applied by some investigators [see, e.g., Tamura et al (1994); Rosenberger (1996)]. In this section, following Sutradhar, and Jowaheer (2006) [see also Sutradhar, Biswas, and Bari (2005)] we discuss a SLPW design to deal with longitudinal count data.

Note that as w_i is the probability of selection of the better treatment for the ith patient, it is convenient to compute w_i by considering two types of balls in an urn, the first type being the indicator for the selection of the better treatment A and the second type for the other treatment. The two types of balls are added to the urn as follows.

(a) As in the beginning we have no reason to believe that any particular treatment is better than the other, we take the initial urn composition in a 50:50 fashion. Thus, the urn will have two types of balls, say α balls of each type at the outset, and the probability that the first patient will be treated by treatment A is 0.5; that is, $\Pr(\delta_1 = 1) = w_1 = 0.5$. For simplicity one may use $\alpha = 1$.

(b) Suppose that at the selection stage of the ith patient $\{y_{rt}\}$ denote all available responses for $r = 1, \ldots, i-1$ and $1 \leq t \leq \min(T, i-r)$. The range of t here depends on the value of r. For example, for the selection time of the ith $(i = 2, \ldots, K)$ patient, $t = 1$ when $r = i-1$. Similarly $t = 1, 2$ for $r = i-2$. Also suppose that at this selection stage we take all these available responses into account and for a suitable τ value and for specific available response y_{rt}, we add τ balls of the same kind by which the patient was treated if $y_{rt} \leq m_0^*$, and add τ balls of the opposite kind in the urn if $y_{rt} > m_0^*$. Here m_0^* is a threshold value of the responses so that any patient with response less than this may be thought to belong to the success group. By the same token, if the response exceeds this threshold value, the patient may be thought to belong to the failure group. Thus, at this stage, we add τ balls for each and every available response. In general τ can be small such as $\tau = 2$, or 4.

(c) On top of the past responses, it may also be sensible to take into account the condition of certain covariates which, along with the treatment (A or B) were responsible for yielding those past responses y_{rt}. For a suitable quantity u_{rt} defined such that a large value of u_{rt} implies the prognostic factor based on a less serious condition of the rth $(r = 1, \ldots, i-1)$ past patient, $G - u_{rt}$ balls of the same kind by which the rth patient was treated and u_{rt} balls of the opposite kind are added, at the treatment selection stage for the ith patient, where $[0,\ G]$ is the domain of u_{rt}.

The above scheme described through **(a)** to **(c)**, produces the selection probabilities $w_i(i = 2, \ldots, K)$ for the cases $2 \leq i \leq T$ as in Exercise 6.4, and for $i > T$ as in Exercise 6.5.

6.8.1.2 Bivariate Random Walk (BRW) Design

Note that in the cross-sectional setup, apart from the randomized play-the-winner rule, there exist some alternative adaptive designs such as the random walk rule [see, e.g., Temple (1981), and Storer (1989)] to collect and analyze the clinical trial data. These random walk rules are variants of the familiar up-and-down rules [Anderson, McCarthy, and Tukey (1946), Derman (1957)]. For example, in the two treatment case, if the $(i-1)$th $(i = 2, \ldots, K,)$ patient is assigned to treatment A, then the ith patient will be assigned to treatment A with probability p_i, and to treatment B with probability q_i, such that $p_i + q_i = 1$. The parameters p_i and q_i depend on the previous patient's response and some random event, such as the result of a biased coin flip.

Remark that in the SLPW design in the previous section, the design weight w_i was mainly dependent on the responses of the individuals $1, 2, \ldots, i-1$, as well as on the conditions of their covariates. Consequently, the construction of any random walk type of rules must be based on past responses as well as covariates. As in the previous section, suppose that a greater value of u_{rt} implies a better condition of the rth past patient and it was a more favorable condition of the patient to treat. By the same token, a smaller value of u_{rt} means that the patient was serious. Now to make sure that this better or serious covariate condition of the past patient does not influence the selection of the treatment for the present ith patient, and also to make sure that the past better response (say, a low value of the response such as $y_{rt} \leq y_0$) gets more weight for the assignment of the patient to the better treatment, one may use a bivariate probability structure given by

$$Pr(u_{rt} \leq u_0, y_{rt} \leq y_0) = p_{rt},\ Pr(u_{rt} \leq u_0, y_{rt} > y_0) = q_{rt},$$

$$Pr(u_{rt} > u_0, y_{rt} \leq y_0) = q_{rt},\ Pr(u_{rt} > u_0, y_{rt} > y_0) = h_{rt},$$

so that $p_{rt} + 2q_{rt} + h_{rt} = 1.0$. Here the parameters are chosen such that $p_{rt} > q_{rt} > h_{rt}$. Note that the bivariate probability structure arises from the consideration of using the past responses and the covariate condition of the patients.

The design weights w_i under this BRW rule are given in Exercise 6.6 for the case $2 \leq i \leq T$, and in Exercise 6.7 for the case $i > T$.

6.8.2 Performance of the SLPW and BRW Designs For Treatment Selection: A Simulation Study

In the last two sections, we have discussed how to construct the longitudinal adaptive design weights represented by w_i for the selection of a better treatment for the ith patient, for all $i = 2, \ldots, K$. We now conduct an empirical study to examine the performance of w_i under both SLPW and BRW designs.

To evaluate w_i under the SLPW design, we use the following steps.

Step 1. Parameter Selection: Clinical Design Parameters

$$\alpha = 1.0, \; ; \; G = 3.0, \;\; \text{and} \;\; \tau = 2 \text{ and } 4.$$

Longitudinal Response Model Parameters

$$K = 100 \text{ subjects}, \;\; p = 3 \text{ covariates}, \;\; \beta_1 = 0.5, 1.00; \;\; \beta_2 = 0.5; \;\; \beta_3 = 0.25,$$

along with Poisson AR(1) responses for $T = 4$ time points with correlation index parameter $\rho = 0.9$. Also, use threshold count $m_0^* = 8$.

Note that the $p = 3$ covariates are denoted by $x_{it} = (\delta_i, x_{it2}, x_{it3})'$. Here δ_i is the treatment selection for the ith patient. Suppose that x_{it2} and x_{it3} are both non-stochastic covariates. Let $x_{it2} = 0, 1, \ldots, 5$ denote the number of chronic diseases for the ith patient at the entry time to the clinical experiment, and $x_{it3} = 1, 2, \ldots, 6$ be the age group of the ith patient. These two covariates are virtually time independent. We generate these covariates as

$$x_{it2} \sim \text{Binomial}(5, p = 0.9)$$

$$z_{it3} \sim \text{Uniform}(20, 80),$$

for all $i = 1, \ldots, K$, and $t = 1, \ldots, T$, and then assign

$$x_{it3} = \begin{cases} 1 & \text{for } 20 \le z_{it3} < 30 \\ 2 & \text{for } 30 \le z_{it3} < 40 \\ 3 & \text{for } 40 \le z_{it3} < 50 \\ 4 & \text{for } 50 \le z_{it3} < 60 \\ 5 & \text{for } 60 \le z_{it3} < 70 \\ 6 & \text{for } 70 \le z_{it3} \le 80. \end{cases}$$

Step 2. Generate Correlated Responses for First Individual: First using $w_1 = \frac{1}{2}$, generate δ_1 such that $Pr[\delta_1 = 1] = w_1$. Now for $i = 1$, that is, for the first patient, use

$$x_{11} = [\delta_1, x_{111}, x_{112}]'$$

and generate y_{11} following

$$y_{11} \sim \text{Poi}(\mu_{11} = \exp(x'_{11}\beta).$$

Next use the stationary Poisson AR(1) model (6.14), that is,

$$y_{1t} = \rho * y_{1,t-1} + d_{11},$$

to generate the remaining three responses, namely $y_{12}, y_{13},$ and y_{14}.

Step 3. Generation of the nonstochastic *u*-Variable: Next to generate w_2, one depends on the y_{11} just generated and also on a u-variable which is a function of the second and third covariates. We now define the nonstochastic u-variable, u_{it}, given by

$$u_{it} = \frac{2}{x_{it2}+1} + \frac{1}{x_{it3}}$$

which ranges from 0.5 to 3. This aids the consideration of $G = 3$ under the SLPW design.

Step 4. Generation of w_i and δ_i for $i = 2, \ldots, K$: Use the formula for w_i from Exercise 6.4 and 6.5. The desired y_{it} values are generated following the model (6.14); that is,

$$y_{1t} = \rho * y_{1,t-1} + d_{1t}. \tag{6.88}$$

Step 5. Generate δ_i. Once w_i is computed, obtain δ_i such that $Pr[\delta_i = 1] = w_i$, and compute $\delta^* = \sum_{i=1}^{K} \delta_i$ in each simulation.

In a manner similar to that of the SLPW design, we now evaluate w_i under the BRW design. To compute w_i in the BRW design, one requires an upper limit for the u-variable, say $u_0 = 1$ and an upper limit for y_{rt}, say $y_0 = 8$ for all past rth individuals at time point $t = 1, \ldots, 4$. By using $\beta_1 = 1.0, \beta_2 = 0.25$, and $\beta_3 = 0$ we generate w_2 and other values of $w_i, i = 3, \ldots, 100$ by using the formulas from Exercise 6.6

and 6.7. For the BRW design we also use $p_{rt} = 0.75, q_{rt} = 0.10$, and $h_{rt} = 0.05$ as the bivariate probabilities depending on the past responses and the values of the u-variable.

Next, in each of 1000 simulations we generate binary values δ_i with corresponding probability w_i, where the w_i are generated as above except that $w_1 = 0.5$. In each simulation we then calculate $\delta^* = \sum_{i=1}^{100} \delta_i$. For different parameter values under two designs, the mean and standard error of δ^* are shown in Table 6.14.

Table 6.14 Simulated mean values and simulated standard errors of the total number of patients $\delta^* = \sum_{i=1}^{100} \delta_i$ receiving the better treatment (A) among $K = 100$ subjects under both SLPW and BRW designs, based on 1000 simulations.

				δ^*	
Design	ρ		β_1	Mean	SE
SLPW	0.9	$\tau = 2$	0.50	62	4.90
			1.00	58	4.73
		$\tau = 4$	0.50	68	4.92
			1.0	61	4.79
BRW	0.9		0.50	61	4.74
			1.00	56	5.01

It is clear from Table 6.14 that the design weights w_i under both SLPW and BRW designs appear to perform well for the selected parameter values. In all cases, the design weights appear to help assign more patients to the better treatment. More specifically, for $\tau = 2$ and $\beta_1 = 0.50$, the SLPW design assigns on the average 62 patients out of 100 to the better treatment A. Similarly for $\beta_1 = 0.50$, the BRW design assigns 61 patients on the average to the better treatment A. Note that all these values of total number of patients receiving treatment A are significant as the standard errors of $\delta^* = \sum_{i=1}^{100} \delta_i$ are reasonably small in all cases. Remark that β_1 in both designs represent the treatment effect. In both SLPW and BRW designs, smaller values of the response variable y indicate that the treatment is better. For example, a fewer number of asthma attacks for an individual implies that the individual received the better treatment. This justification also follows, for example, from the formulas for w_i in Exercises 6.4 and 6.6. This is because as the threshold point m_0^* in the SLPW design and the cut point (y_0, u_0) in the BRW design are predetermined and fixed, the smaller values of the response variable y will produce many of $I(y_{rt}) \leq m_0^*$ as 1 in the formula for w_i in Exercise 6.4, and $\delta_{y_{rt}} p_{rt}$ in the formula for w_i in Exercise 6.6 will contribute significantly. Thus, the better treatment should produce smaller values of y in the present setup. This in turn means that the smaller values of β_1 should indicate the better treatment. Consequently, the formulation of the design weights for both SLPW and BRW designs appear to work well as more patients are seen to be assigned to treatment A when $\beta_1 = 0.5$ as compared to $\beta_1 = 1.0$.

6.8.3 Weighted GQL Estimation for Treatment Effects and Other Regression Parameters

In previous sections, the repeated count responses for the ith individual were represented by a vector $y_i = [y_{i1}, \ldots, y_{it}, \ldots, y_{iT}]'$ with its mean vector μ_i, and covariance matrix $\Sigma_i^*(\rho) = A_i^{1/2} C_i^*(\rho) A_i^{1/2}$ (6.26) under the stationary correlation models or $\Sigma_i^{(ns)}(\rho) = A_i^{1/2} C_i^{(ns)}(x_i, \rho) A_i^{1/2}$ (6.56) under the nonstationary correlation models. It was, however, demonstrated in Sections 6.8.1 and 6.8.2 under the longitudinal adaptive clinical trial setup, that a treatment is selected first for the ith individual based on adaptive design weight w_i, and then the responses are collected. To reflect this operation, we now denote the response vector as

$$y_i(w_i) = [y_{i1}(w_i), \ldots, y_{it}(w_i), \ldots, y_{iT}(w_i)]'$$

and its mean vector and stationary correlations based covariance matrix, for example, by

$$\mu_i(w_{i0}), \text{ and } \Sigma_i^*(w_{i0}, \rho),$$

respectively, where w_{i0} is the limiting value of w_i, for example, $w_{i0} = E[w_i]$.

6.8.3.1 Formulas for $\mu_i(w_{i0})$, and $\Sigma_i^*(w_{i0}, \rho)$:

Construction of the Mean Vector $\mu_i(w_{i0})$ Let

$$z'_{it} = x'_{it}|_{\delta_i=1} = (1, x^{*'}_{it}), \text{ and } z^{*'}_{it} = x'_{it}|_{\delta_i=0} = (0, x^{*'}_{it}),$$

where $x^*_{it} = (x_{it2}, \ldots, x_{itp})'$. Also, define

$$\mu^*_{rt1} = \exp(z'_{rt}\beta), \text{ and } \mu^*_{rt2} = \exp(z^{*'}_{rt}\beta). \tag{6.89}$$

Now by taking the average over the distribution of δ_i, it follows from (6.85) that the unconditional mean of Y_{it}, that is, $\mu_{it}(w_{i0})$ has the formula given by

$$\begin{aligned} E(Y_{it}|x^*_{it}) &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\delta_2,\ldots,\delta_{i-1}} E(Y_{it}|\delta_i, \ldots, \delta_1) \\ &= w_{i0} \exp(z'_{it}\beta) + (1 - w_{i0}) \exp(z^{*'}_{it}\beta) \\ &= w_{i0}\mu^*_{it1} + (1 - w_{i0})\mu^*_{it2} \\ &= \mu_{it}(w_{i0}), \end{aligned} \tag{6.90}$$

where for $i = 1, \ldots, K$, w_{i0} is the expectation of w_i, with $w_i = Pr(\delta_i = 1|y_{H_{i-1}})$ as defined in Exercises 6.4 and 6.5 for the SLPW design, and in Exercises 6.6 and 6.7, for the BRW design. More specifically, for the SLPW design, w_{i0} can be computed for the case $2 \leq i \leq T$ as

$$
\begin{aligned}
w_{i0} &= E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_i|\delta_1,\delta_2,\ldots,\delta_{i-1}} E(\delta_i|y_{H_{i-1}}) \\
&= \frac{1}{2\alpha + \frac{1}{2}i(i-1)(G+\tau)} \\
&\quad \times \left[\alpha + \sum_{r=1}^{i-1} \sum_{t=1}^{i-r} [\{(G-u_{rt}) + \tilde{\mu}_{rt1}\tau\} w_r \right. \\
&\quad \left. + \{u_{rt} + (1-\tilde{\mu}_{rt2})\tau\}(1-w_r)] \right], \qquad (6.91)
\end{aligned}
$$

and for the case $i > T$ as

$$
\begin{aligned}
w_{i0} &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\delta_2,\cdots,\delta_{i-1}} E(\delta_i|y_{H_{i-1}}) \\
&= \left\{ 2\alpha + (G+\tau)T \left(i - \frac{T+1}{2} \right) \right\}^{-1} \\
&\quad \times \left[\alpha + \sum_{r=1}^{i-T} \sum_{t=1}^{T} \{(G-u_{rt}+\tilde{\mu}_{rt1}\tau)w_r + (u_{rt}+(1-\tilde{\mu}_{rt2})\tau)(1-w_r)\} \right. \\
&\quad + \sum_{r=i-T+1}^{i-1} \sum_{t=1}^{i-r} \{((G-u_{rt}) + \tilde{\mu}_{rt1}\tau)w_r \\
&\quad \left. + (u_{rt} + (1-\tilde{\mu}_{rt2})\tau)(1-w_r)\} \right], \qquad (6.92)
\end{aligned}
$$

with

$$\tilde{\mu}_{rt1} = \int_0^{m_0^*} f(y_{rt}|\theta_{rt} = z_{rt}'\beta) = \sum_{k=0}^{m_0^*} \frac{\exp(-\mu_{rt1}^*)(\mu_{rt1}^*)^k}{k!}$$

and

$$\tilde{\mu}_{rt2} = \int_0^{m_0^*} f(y_{rt}|\theta_{rt} = z_{rt}^{*\prime}\beta) = \sum_{k=0}^{m_0^*} \frac{\exp(-\mu_{rt2}^*)(\mu_{rt2}^*)^k}{k!},$$

where m_0^* is the threshold count as mentioned before.

Note that the computation of the unconditional mean vector $\mu_i(w_{i0})$ for the BRW design is similar to that of SLPW design, and hence omitted.

Construction of the Covariance Matrix $\Sigma_i^*(w_{i0}, \rho)$

Next, we construct the unconditional covariance matrix $\Sigma_i^*(\rho)$ of the Y_i vector as follows. Recall that given $\delta_1, \delta_2, \ldots, \delta_i$, or simply say, given δ_i, the conditional variance of Y_{it} and the conditional covariance between Y_{it} and Y_{iv} are given in (6.85) and (6.87), respectively. Now by similar arguments as for the construction of the mean vector, the unconditional covariance between Y_{it} and Y_{iv} may be computed as

$$\begin{aligned}
\text{cov}[(Y_{it},Y_{iv})|x^*_{it},x^*_{iv}] &= E_{\delta_1}E_{\delta_2|\delta_1}\ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}[\text{cov}(Y_{it},Y_{iv})|\delta_i] \\
&\quad + \text{cov}_{\delta_1,\ldots,\delta_i}[E(y_{it}|\delta_i),E(y_{iv}|\delta_i)], \\
&= E_{\delta_1}E_{\delta_2|\delta_1}\ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}[\rho_{|t-v|}\{\exp[(\theta_{it}+\theta_{iv})'\beta]\}^{1/2}] \\
&\quad +\text{cov}_{\delta_1,\ldots,\delta_i}\{\exp[(\theta_{it}+\theta_{iv})'\beta]\} \\
&= \rho_{|t-v|}\left[w_{i0}\{\mu^*_{it1}\mu^*_{iv1}\}^{1/2}+(1-w_{i0})\{\mu^*_{it2}\mu^*_{iv2}\}^{1/2}\right] \\
&\quad +w_{i0}\{\mu^*_{it1}\mu^*_{iv1}\}+(1-w_{i0})\{\mu^*_{it2}\mu^*_{iv2}\}-\mu_{it}(w_{i0}\mu_{iv}(w_{i0}) \\
&= \sigma^*_{ijk}(w_{i0},\rho), \text{ say,} \qquad (6.93)
\end{aligned}$$

where μ^*_{it1} and μ^*_{it2} are given as in (6.89), and $\mu_{it}(w_{i0})$ is given as in (6.90). For $t=v$, equation (6.93) yields the unconditional variance of y_{it} given by

$$\text{var}(Y_{it}|x^*_{it}) = \mu^*_{it}+\{w_{i0}\mu^{*2}_{it1}+(1-w_{i0})\mu^{*2}_{it2}\}-\mu^{*2}_{it}. \qquad (6.94)$$

The construction of the covariance matrix $\Sigma^*_i(w_{i0},\rho)=(\sigma^*_{ijk}(w_{i0},\rho))$, say, is now completed by (6.93) and (6.94).

6.8.3.2 Weighted GQL Estimation of β

Note that $\beta=[\beta_1,\beta_2,\ldots,\beta_p]'$ is the effect of the covariate

$$x_{it}=[\delta_i,x^{*'}_i t]'=[\delta_i,x_{it2},\ldots,x_{itp}]'$$

on y_{it} for all $i=1,\ldots,K$, and $t=1,\ldots,T$, where y_{it} is now collected based on longitudinal adaptive design scheme and is represented by $y_{it}(w_i)$. Because $E[Y_i(w_i)]=\mu_i(w_{i0})$ by (6.90), and $\text{var}[Y_i(w_i)]=\Sigma^*_i(w_{i0},\rho)$ by (6.93) and (6.94), similar to the construction of the GQL estimating equation (6.26) or (6.56), we may now construct a weighted GQL estimating equation for β given by

$$\sum_{i=1}^{K}\frac{\partial\mu'_i(w_{i0})}{\partial\beta}\Sigma^{*-1}_i(w_{i0},\hat{\rho})(y_i(w_i)-\mu_i(w_{i0})=0. \qquad (6.95)$$

where $\hat{\rho}$ is a consistent estimate of ρ, the longitudinal correlation index parameter of the model. Now, by treating the data as though they follow the stationary correlation structure, one may apply the MM and equate the sample auto-covariance to the autocovariance of the data given by (6.93) and obtain a moment estimate of ρ_ℓ ($\ell=|t-v|=1,\ldots,T-1$) as

$$\hat{\rho}_\ell=\frac{N_1-N_2}{D}, \qquad (6.96)$$

where

$$N_1 = \frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}[(y_{it}-\mu_{it}(w_{i0}))(y_{iv}-\mu_{iv}(w_{i0}))/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it}-\mu_{it}(w_{i0})]^2/KT}$$

$$N_2 = -\frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}[w_{i0}\mu^*_{it1}\mu^*_{iv1}+(1-w_{i0})\mu^*_{it2}\mu^*_{iv2}-\mu_{it}(w_{i0})\mu_{iv}(w_{i0})]/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[\mu_{it}(w_{i0})-\mu^2_{it}(w_{i0})+w_{i0}\mu^{*2}_{it1}+(1-w_{i0})\mu^{*2}_{it2}]/KT},$$

and

$$D = \frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}\left[w_{i0}\{\mu^*_{it1}\mu^*_{iv1}\}^{1/2}+(1-w_{i0})\{\mu^*_{it2}\mu^*_{iv2}\}^{1/2}\right]/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[\mu_{it}(w_{i0})-\mu^2_{it}(w_{i0})+w_{i0}\mu^{*2}_{it1}+(1-w_{i0})\mu^{*2}_{it2}]/KT}.$$

For given $\hat{\rho}_\ell$ (a function of $\hat{\rho}$), the solution of (6.95) may easily be obtained by using the Newton−Rapson iterative equation.

$$\hat{\beta}_{(m+1)} = \hat{\beta}_{(m)} + \left[\sum_{i=1}^{K}\frac{\partial \mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})\frac{\partial\mu_i(w_{i0})}{\partial\beta'}\right]^{-1}_m \times\left[\sum_{i=1}^{K}\frac{\partial \mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})(y_i(w_i)-\mu_i(w_{i0}))\right]_m, \quad (6.97)$$

where $\hat{\beta}_{(m)}$ is the value of β at the mth iteration and $[\cdot]_m$ denotes that the expression within brackets is evaluated at $\hat{\beta}_{(m)}$. Let $\hat{\beta}_{WGQL}$ be the solution of (6.97), which is consistent for β.

Under some mild regularity conditions, it may be shown from (6.97) that for large K, $\hat{\beta}_{WGQL}$ has an asymptotically p-dimensional normal distribution with mean β and covariance matrix var$(\hat{\beta}_{WGQL})$ which may be consistently estimated by using the sandwich type estimator given by

$$\hat{\text{var}}(\hat{\beta}_{WGQL}) = \left[\sum_{i=1}^{K}\frac{\partial \mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})\frac{\partial\mu_i(w_{i0})}{\partial\beta'}\right]^{-1}$$

$$+\left[\sum_{i=1}^{K}\frac{\partial \mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})\frac{\partial\mu_i(w_{i0})}{\partial\beta'}\right]^{-1}$$

$$\times\left[2\sum_{i<r}^{K}\frac{\partial \mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})(y_i-\mu_i(w_{i0}))\right.$$

$$\left.\times\,(y_r-\mu_r(w_{i0}))'\Sigma_r^{*-1}(w_{r0},\hat{\rho})\frac{\partial\mu_r(w_{r0})}{\partial\beta'}\right]$$

$$\times \left[\sum_{i=1}^{K} \frac{\partial \mu_i'(w_{i0})}{\partial \beta} \Sigma_i^{*-1}(w_{i0}, \hat{\rho}) \frac{\partial \mu_i(w_{i0})}{\partial \beta'} \right]^{-1}. \tag{6.98}$$

Formula for the Derivative $(\partial \mu_i'(w_{i0}))/\partial \beta$ in (6.95)

As

$$\frac{\partial \mu_{it}(w_{i0})}{\partial \beta} = w_{i0}\mu_{it1}^* z_{it} + (1 - w_{i0})\mu_{it2}^* z_{it}^*,$$

the $p \times T$ matrix $\partial \mu_i'(w_{i0})/\partial \beta$ is computed as

$$\frac{\partial \mu_i'(w_{i0})}{\partial \beta} = w_{i0} Z_i' A_{i1} + (1 - w_{i0}) Z_i^{*'} A_{i2}, \tag{6.99}$$

where $Z_i' = (z_{i1}, \ldots, z_{it}, \ldots, z_{iT})$ and $Z_i^{*'} = (z_{i1}^*, \ldots, z_{it}^*, \ldots, z_{iT}^*)$ are $p \times T$ matrices, $A_{i1} = \text{diag}[\mu_{i11}^*, \ldots, \mu_{iT1}^*]$, and $A_{i2} = \text{diag}[\mu_{i12}^*, \ldots, \mu_{iT2}^*]$, with

$$\mu_{it1}^* = \exp(z_{it}'\beta), \;\; \mu_{it2}^* = \exp(z_{it}^{*'}\beta),$$

where $z_{it} = (1, x_{it}^{*'})'$ and $z_{it}^* = (0, x_{it}^{*'})'$, for all $t = 1, \ldots, T$.

Exercises

6.1. (Section 6.5.1.1) [Likelihood estimation for nonstationary AR(1) model]
Consider the nonstationary AR(1) model given by (6.44). Then demonstrate that similar to that (6.23) of the stationary AR(1) model (6.14), one may write the likelihood function for the model (6.44) as

$$L(\beta, \rho) = \Pi_{i=1}^{K} [f(y_{i1}) \Pi_{t=2}^{T} f(y_{it}|y_{i,t-1})],$$

with

$$f(y_{it}|y_{i,t-1}) = \exp[-(\mu_{it} - \rho\mu_{i,t-1})]$$
$$\times \sum_{s=1}^{\min(y_{it}, y_{i,t-1})} \frac{(y_{i,t-1})!\rho^s(1-\rho)^{y_{i,t-1}-s}(\mu_{it} - \rho\mu_{i,t-1})^{y_{it}-s}}{s!(y_{i,t-1} - s)!(y_{it} - s)!}.$$

Now, argue that the likelihood estimation of β and ρ, is extremely complicated.

6.2. (Section 6.5.1.1) [Conditional moments for nonstationary AR(1) model]
Show either by using the conditional density from Exercise 6.1, or by direct computation from the model (6.44), that for $t = 2, \ldots, T$, the conditional mean and variance of y_{it} given $y_{i,t-1}$ have the formulas:

$$E[Y_{it}|y_{i,t-1}] = \mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1})$$
$$\text{var}[Y_{it}|y_{i,t-1}] = \mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1}) - \rho^2 y_{i,t-1}.$$

Next, verify that for $u < t$, the conditional covariance has the formula

$$\text{cov}[\{Y_{iu}, Y_{it}\}|y_{i,u-1}, y_{i,t-1}] = 0.$$

6.3. (Section 6.5.2) [Conditional GQL estimating equation]
Denote the conditional mean and the variance in Exercise (6.2) by $\mu^*_{it|t-1}$ and $\lambda_{itt|t-1}$, respectively. Let $\mu^* = [\mu_{i1}, \mu^*_{i2|1}, \ldots, \mu^*_{it|t-1}, \ldots, \mu^*_{iT|T-1}]'$ be the $T \times 1$ conditional mean vector, and $\Lambda_i = \text{diag}[\mu_{i1}, \lambda_{i22|1}, \ldots, \lambda_{itt|t-1}, \ldots, \lambda_{iTT|T-1}]$ is the $T \times T$ conditional covariance matrix of y_i. Then, similar to (6.56), argue that a consistent estimator of β can also be obtained by solving the conditional GQL estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \mu^{*\prime}}{\partial \beta} \Lambda_i^{-1}(\hat{\rho})(y_i - \mu^*) = 0,$$

where $\hat{\rho}$ is obtained by using (6.58) as in the unconditional estimation. Also, derive the formulas for the elements of the $p \times T$ derivative matrix $\partial \mu^{*\prime}/\partial \beta$. Comment on the relative efficiency of this conditional GQL estimator of β as compared to the unconditional GQL estimator obtained from (6.56).

6.4. (Section 6.8.1.1) [w_i for the case $2 \le i \le T$ under **SLPW** rule]
As the selection of the ith patient is made at the ith time point, by this time, the $(i-1)$th patient has yielded one response and $(i-2)$th patient has yielded two responses and so on. Use the rules **(a), (b),** and **(c)** from the Section 6.8.1.1 and argue that at this treatment selection stage for the ith patient, there are

$$n^*_{i-1} = 2\alpha + \sum_{r=1}^{i-1} \sum_{t=1}^{i-r} (G + \tau) = 2\alpha + \frac{1}{2} i(i-1)(G + \tau)$$

balls in total in the urn. Also justify that among these balls, there are

$$n^*_{i-1,1}(y_{H_{i-1}}) = \alpha + \sum_{r=1}^{i-1} \sum_{t=1}^{i-r} [\delta_r\{(G - u_{rt}) + I[y_{rt} \le m^*_0]\tau\}$$
$$+ (1 - \delta_r)\{u_{rj} + I[y_{rt} > m^*_0]\tau\}]$$

balls of first type, where $y_{H_{i-1}}$ indicates the history of responses from the past $i-1$ patients. The number of second type of balls may be denoted by $n^*_{i-1,2}(y_{H_{i-1}})$. It then follows that for given $y_{H_{i-1}}$, the conditional probability that $\delta_i = 1$ is given by

$$w_i = Pr(\delta_i = 1|y_{H_{i-1}}) = n^*_{i-1,1}(y_{H_{i-1}})/n^*_{i-1}.$$

6.5. (Section 6.8.1.1) [w_i for the case $i > T$ under **SLPW** rule]
Argue that under this case, at the treatment selection stage for the ith patient, there

are

$$\tilde{n}_{i-1} = 2\alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}(G+\tau) + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}(G+\tau)$$

balls in total in the urn. Also argue that among these balls, there are $\tilde{n}_{i-1,1}(y_{H_{i-1}})$ balls of first type, where

$$\begin{aligned}\tilde{n}_{i-1,1}(y_{H_{i-1}}) = \alpha &+ \sum_{r=1}^{i-T}\sum_{t=1}^{T}[\delta_r\{(G-u_{rt})+I[y_{rt}\le m_0^*]\tau\} \\ &+(1-\delta_r)\{u_{rt}+I[y_{rt}>m_0^*]\tau\}] \\ &+ \sum_{r=1-T+1}^{i-1}\sum_{t=1}^{i-r}[\delta_r\{(G-u_{rt})+I[y_{rt}\le m_0^*]\tau\} \\ &+(1-\delta_r)\{u_{rt}+I[y_{rt}>m_0^*]\tau\}].\end{aligned}$$

Clearly, for this $i > T$ case, one may then evaluate the design weight w_i as

$$w_i = \frac{\tilde{n}_{i-1,1}(y_{H_{i-1}})}{\tilde{n}_{i-1}}.$$

6.6. (Section 6.8.1.2) [w_i for the case $2 \le i \le T$ under **BRW** rule]
Let $\delta_{u_{rt}} = 1$ for $u_{rt} \le u_0$ and $\delta_{u_{rt}} = 0$ otherwise. Similarly, let $\delta_{y_{rt}} = 1$ for $y_{rt} \le y_0$ and $\delta_{y_{rt}} = 0$ otherwise. Verify, in the fashion similar to that of Exercise 6.4 that under the BRW rule, the design weight w_i has the formula

$$w_i = \frac{\sum_{r=1}^{i-1}\sum_{t=1}^{i-r}[\delta_{u_{rt}}g(y_{rt})]+[(1-\delta_{u_{rt}})s(y_{rt})]}{\sum_{r=1}^{i-1}\sum_{t=1}^{i-r}(p_{rt}+2q_{rt}+h_{rt})},$$

where $g(y_{rt}) = \delta_{y_{rt}}p_{rt} + (1-\delta_{y_{rt}})q_{rt}$, and $s(y_{rt}) = \delta_{y_{rt}}q_{rt} + (1-\delta_{y_{rt}})h_{rt}$.

6.7. (Section 6.8.1.2) [w_i for the case $i > T$ under **BRW** rule]
For this case, make an argument similar to that of Exercise 6.5 for the SLPW design, and justify under the BRW rule, that w_i has the formula given by

$$\begin{aligned}w_i = &\frac{1}{0.5i(i-1)-0.5(i-T)(i-T-1)} \\ &\times[\sum_{r=1}^{i-T}\sum_{t=1}^{T}[\delta_{u_{rt}}g(y_{rt})]+[(1-\delta_{u_{rt}})s(y_{rt})] \\ &+\sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}[\delta_{u_{rt}}g(y_{rt})]+[(1-\delta_{u_{rt}})s(y_{rt})]],\end{aligned}$$

where $g(y_{rt})$ and $s(y_{rt})$ are defined as in Exercise 6.6.

References

1. Amemiya, T. (1985). *Advanced Econometrics.* Cambridge, MA: Harvard University Press.
2. Anderson, T. W., McCarthy, P. J., & Tukey, J. W. (1946). Staircase Method of Sensitivity Testing. *Naval Ordinance Report*, 35 – 46. Princeton, NJ: Statistical Research Group.
3. Atkinson, A. C. (1999). Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statist. Med.,* 18, 1741 – 1752.
4. Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika,* 82, 407 – 410.
5. Derman, C. (1957). Nonparametric up and down experimentation. *Ann. Math. Stat.*, 28, 795 – 798.
6. Durham, S. D. & Flournoy, N. (1994). Random walks for quantile estimation. In *Statistical Decision Theory and Related Topics V* (S. S. Gupta and J. O. Berger, eds.) 467 – 476. New York: Springer.
7. Farewell, V. T. , Viveros, R. , & Sprott, D. A. (1993). Statistical consequences of an adaptive treatment allocation in a clinical trial. *Canad. J. Statist.,* 21, 21 – 27.
8. Jennison, C. & Turnbull, B. W. (2001). Group sequential tests with outcome-dependent treatment assignment. *Sequential Anal.,* 20, 209 – 234.
9. Liang, K. Y. & Zeger, S . L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika,* 78, 13 – 22.
10. Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.
11. McCullagh, P. (1983). Quasilikelihood functions. *Ann. Statist.* 11, 59 – 67.
12. McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advan. Appl. Probab.,* 20, 822 – 835.
13. Nelder, J. & Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. A135,* 370 – 384.
14. Pocock, S. J. & Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics,* 31, 103 – 115.
15. Rosenberger, W. F. (1996). New directions in adaptive designs. *Statist. Sci.,* 11, 137 – 149.
16. Royall, R. M. (1991). Ethics and statistics in randomized clinical trials (with discussion). *Statist. Sci.,* 6, 52 – 62.
17. Smith, R. L. S. (1984). Properties of bias coin designs in sequential clinical trials.*Annal. Statist.,* 12, 1018 – 1034.
18. Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statist. Sci.,* 18, 377 – 393.
19. Sutradhar, B.C., Biswas, A., & Bari, W. (2005). Marginal regression for binary longitudinal data in adaptive clinical trials. *Scand. J. Statist.,* 32, 93 – 113.
20. Sutradhar, B. C. & Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika,* 86, 459 – 465.
21. Sutradhar, B. C. & Jowaheer, V. (2006). Analyzing longitudinal count data from adaptive clinical trials: a weighted generalized quasi-likelihood approach. *J. Statist. Comput. Simul.,* 76, 1079 – 1093.
22. Sutradhar, B. C. & Kovacevic, M. (2000). Analyzing ordinal longitudinal survey data: Generalized estimating equations approach. *Biometrika,* 87, 837 – 848.
23. Storer, B, E, (1989). Design and analysis of phase 1 clinical trial. *Biometrics,*, 45, 925 – 937.
24. Tamura, R. N. , Faries, D. E. , Andersen, J. S., & Heiligenstein, J. H. (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *J. Amer. Stat. Assoc.,* 89, 768 – 776.
25. Temple, R. (1981). Government view points of clinical trials. *Drug Inf. J.,* 16, 10 – 17.
26. Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika,* 61, 439 – 447.
27. Wei, L. J. & Durham, S. (1978). The randomized play-the-winner rule in medical trials. *J. Am. Statist. Assoc.,* 73, 840 – 843.

28. Wei, L. J., Smythe, R. T. Lin, D. Y., & Park, T. S. (1990). Statistical inference with data-dependent treatment allocation rules. *J. Am. Statist. Assoc.,* 85, 156 – 62.
29. Zelen, M. (1969). Play-the-winner rule and the controlled clinical trial. *J. Am. Statist. Assoc.,* 64, 131 – 46.

Appendix

Table 6A. Health care utilization data for six years from 1985 to 1990 collected by Health Science Center, Memorial University, St. John's, Canada. [Code: column 1 (C1)-Family identification; C2-Member identification; C3-Gender (1 for male, 2 for female); C4-Chronic disease status (0 for no chronic disease, 1 for 1 chronic disease and so on); C5-Education level (1 for less than high school, 2 for high school, 3 for university graduate, and 4 for post graduate); C6-Age at 1985; C7-C12-Number of physician visits from 1985 to 1990]

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
7	101	2	3	3	51.7	10	15	10	6	7	14
7	102	1	2	2	55.4	0	6	0	4	2	0
7	203	2	1	4	24.9	12	6	2	0	3	2
7	204	2	0	4	21.5	0	1	1	0	0	0
27	101	2	1	3	49.5	2	11	8	7	7	3
27	102	1	1	4	50.7	13	13	16	12	18	12
27	203	1	0	4	20.2	1	5	0	2	0	0
27	203	1	0	4	20.2	2	3	7	1	0	0
36	101	2	2	3	49.7	5	5	4	18	11	9
36	102	1	1	3	54.6	1	0	0	2	1	1
36	203	2	0	3	26.0	10	6	9	9	21	16
36	204	1	0	2	22.4	3	4	1	0	4	1
189	101	2	1	3	58.6	4	3	1	3	0	6
189	102	1	0	4	58.3	1	0	0	3	0	3
189	203	2	3	2	31.7	8	4	4	12	12	7
189	204	2	1	3	20.2	2	0	6	2	2	5
436	101	2	0	1	62.1	10	8	7	10	8	11
436	102	1	0	1	68.9	6	5	2	6	4	6
436	203	1	0	3	31.8	1	3	4	0	0	0
436	204	1	0	4	23.8	2	2	5	0	0	0
469	101	2	4	2	44.1	4	1	6	7	13	3
469	102	1	0	3	47.5	2	0	1	0	1	1
469	203	1	0	3	23.7	2	4	3	2	1	0
469	204	1	2	4	21.2	5	5	5	0	8	0
574	101	2	0	1	47.2	4	10	12	17	13	10
574	102	1	4	1	52.9	8	9	14	23	22	15
574	203	2	1	3	23.2	5	3	6	6	5	7
574	204	1	0	2	21.9	2	0	3	3	1	1
580	101	2	2	1	41.9	2	5	1	0	1	0
580	102	1	0	2	44.2	1	1	4	24	5	2
580	203	2	1	2	20.5	13	11	11	16	18	21
580	204	2	0	2	23	9	3	4	3	19	3
706	101	2	2	3	40.7	17	5	1	5	3	2
706	102	1	0	1	42.9	1	1	7	6	1	0
706	203	1	0	3	21.5	1	3	0	3	0	0
706	204	1	0	3	19.9	0	0	0	0	0	0

Table Cont'd

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
754	101	2	1	2	49.8	8	2	5	12	8	8
754	102	1	0	1	50.8	0	2	0	0	2	0
754	203	1	1	2	21.3	0	0	0	0	1	0
754	204	1	0	4	25.3	1	1	1	0	0	2
758	101	2	1	2	60.9	2	5	1	1	0	0
758	102	1	0	4	63.7	1	0	0	0	0	0
758	203	1	1	4	22.8	0	0	0	2	1	0
758	204	1	0	4	20.9	2	11	4	11	10	4
921	101	2	1	1	50.8	0	3	0	3	7	14
921	202	1	1	1	26.4	1	1	4	3	5	2
921	203	2	1	3	25.2	3	2	2	1	2	2
921	204	2	0	2	21.9	3	2	4	2	5	16
965	101	2	1	1	44.8	13	18	13	13	15	17
965	102	1	2	1	48.6	4	2	0	3	0	6
965	203	1	0	3	25	4	3	1	0	6	2
965	204	1	0	3	20.9	2	3	1	1	3	1
993	101	2	3	1	67.3	2	3	3	2	4	3
993	203	1	2	1	31.3	2	0	1	1	2	3
993	204	2	1	2	22	11	6	3	4	17	8
993	205	1	0	1	22.3	1	1	4	9	4	1
1054	101	2	0	2	41.1	1	11	3	5	2 4	9
1054	102	1	2	1	43.6	3	4	10	4	11	11
1054	203	2	1	4	22.2	4	2	3	4	14	11
1054	204	2	2	4	20.3	1	4	3	5	10	9
1120	101	2	3	1	52.7	2	9	2	1	9	7
1120	102	1	0	1	63.1	0	0	0	0	0	0
1120	203	2	0	4	32.2	12	7	27	11	5	13
1120	204	1	1	2	26	1	3	0	3	10	3
1269	101	2	0	4	56.1	1	3	1	9	10	14
1269	102	1	1	4	56.3	4	0	3	8	4	4
1269	203	1	0	4	22	2	0	2	0	2	1
1269	204	2	0	4	20.5	0	0	0	1	0	0
1333	101	2	1	1	50.9	2	2	1	0	0	0
1333	102	1	0	1	49.5	3	6	2	9	5	4
1333	203	2	0	3	22.6	0	0	0	0	0	0
1333	204	1	0	2	20.6	0	0	0	1	4	12
1344	101	2	2	1	46.4	0	0	0	2	2	3
1344	203	1	0	1	24	0	1	0	0	0	0
1344	204	1	0	1	28.8	0	0	0	0	0	0
1344	205	1	1	1	20.3	2	0	1	1	0	1
1361	101	2	0	1	71.6	4	7	9	8	3	8
1361	202	2	0	3	35.3	2	4	7	9	10	6

Table Cont'd

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
1361	203	1	0	2	33	3	3	5	2	0	3
1361	204	1	0	3	27.4	1	1	2	2	2	3
1397	101	2	0	3	25.3	7	3	5	7	5	5
1397	102	1	1	1	53	2	4	5	6	6	3
1397	203	1	0	4	27.3	2	0	0	0	0	0
1397	204	1	0	3	22	12	1	2	2	4	4
1637	101	2	1	4	43.5	6	10	2	2	3	3
1637	102	1	1	4	47.4	0	3	4	1	0	0
1637	203	1	0	4	23.1	0	0	0	1	1	0
1637	204	1	1	4	21.7	1	2	2	4	5	2
1664	101	2	2	4	47.2	25	9	8	14	12	29
1664	102	1	2	2	49.2	4	3	9	0	10	4
1664	203	2	0	4	23.5	3	3	0	2	2	1
1664	204	1	1	4	22.3	1	1	0	0	0	0
1669	101	2	0	2	50.6	0	0	0	2	4	1
1669	202	2	0	3	24.7	7	5	5	12	7	6
1669	203	1	0	4	22.5	0	0	1	1	2	0
1669	204	1	0	2	20.9	0	0	1	0	0	3
1682	101	2	1	1	62.1	0	2	3	1	0	0
1682	102	1	4	1	65.2	7	0	0	0	0	0
1682	203	1	3	3	29	9	9	12	5	4	4
1682	404	2	4	1	74.9	13	17	16	15	14	10
1702	101	2	2	1	59.2	6	5	2	1	1	6
1702	102	1	2	1	64	0	0	0	0	0	2
1702	203	1	1	1	21.1	0	0	0	0	0	0
1702	304	2	3	1	85.2	6	7	8	6	24	0
1703	101	2	1	3	56.9	3	4	3	10	4	14
1703	202	1	0	4	25.5	0	0	0	0	0	0
1703	204	2	0	4	22.1	1	0	1	3	0	0
1703	305	2	1	2	80.5	5	7	4	8	4	8
1728	101	2	1	1	40.1	5	3	2	2	2	1
1728	102	1	4	3	51.5	12	13	10	7	22	19
1728	203	2	1	2	24.3	10	11	4	5	7	3
1728	204	1	0	3	20.4	3	2	3	2	2	2
1737	101	2	3	2	43.8	11	6	9	4	4	4
1737	102	1	1	4	44.1	6	0	8	1	0	8
1737	203	2	0	3	21.9	1	4	10	8	25	10
1737	204	1	0	4	22.9	0	0	0	0	0	0
1751	101	2	5	2	52	9	12	11	6	18	15
1751	102	1	0	1	55.5	0	0	2	0	1	0
1751	203	1	1	1	23.6	3	2	8	2	3	6
1751	204	1	0	1	22.6	1	8	3	2	1	3

Table Cont'd

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
1838	101	2	0	2	44.7	3	3	3	2	10	11
1838	102	1	1	1	46	3	1	2	2	0	3
1838	203	1	0	4	23.5	2	3	1	4	1	0
1838	404	2	1	1	76.4	0	0	7	5	8	4
1876	101	2	1	1	46.7	0	0	4	4	0	2
1876	102	1	1	3	51.1	2	10	10	16	10	6
1876	203	2	0	3	24.6	5	2	0	0	0	0
1876	205	2	4	4	21	2	1	1	2	3	5
1925	101	2	1	3	52.6	19	4	12	9	7	5
1925	102	1	0	2	60.2	4	15	13	5	1	7
1925	203	2	0	4	21.5	9	6	4	13	8	0
1925	204	1	0	4	23.2	0	0	1	0	0	0
1935	101	2	1	3	65.9	2	1	3	4	5	12
1935	102	1	1	1	67.6	9	6	7	8	7	7
1935	203	1	0	2	25.6	2	1	0	0	0	0
1935	204	2	0	3	38.4	4	2	4	9	17	18
2046	101	2	0	1	56.3	11	17	4	3	12	9
2046	202	1	0	1	33.4	0	0	0	0	0	0
2046	203	1	0	2	27.8	1	1	0	3	3	9
2046	204	2	0	3	25	0	3	4	5	5	8
2076	101	2	2	3	52	5	3	6	8	3	3
2076	102	1	1	1	53.8	2	0	3	7	6	2
2076	203	2	0	4	24.6	14	11	5	1	2	0
2076	204	1	3	3	31.4	2	1	4	3	4	14
41	102	1	0	1	54	0	0	0	0	0	0
41	203	2	0	4	22	2	2	2	9	7	0
41	204	1	0	4	23	3	2	2	4	7	0
101	101	2	1	1	62.8	2	0	0	0	1	0
101	102	1	5	1	65.9	2	2	5	10	7	2
101	203	1	1	3	24.2	0	0	0	0	0	0
129	101	2	3	1	56.3	10	14	7	9	9	13
129	102	1	1	1	57.1	9	15	8	10	13	2
129	204	1	0	4	21.6	1	1	4	1	0	0
208	102	1	0	4	50.5	0	0	0	7	11	12
208	203	1	0	4	25.3	0	1	1	1	4	1
208	204	1	0	3	23.8	1	1	1	1	0	1
219	101	2	4	1	62.5	11	17	8	18	23	17
219	203	2	1	1	40.4	9	4	2	6	4	2
219	204	2	1	4	21.3	5	2	1	4	0	0

Table Cont'd

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
522	102	1	0	1	51.2	1	5	7	7	9	6
522	203	2	1	3	21.6	11	7	3	8	20	19
522	204	2	1	2	24.4	12	7	19	6	12	7
605	101	2	1	1	58.2	2	6	0	2	0	2
605	102	1	0	1	58.6	0	0	0	0	0	1
605	203	1	1	2	21.3	0	0	0	0	1	2
622	203	1	0	1	25	0	0	0	0	0	0
622	204	2	0	1	30.5	0	0	0	0	0	0
622	205	2	0	1	22.4	3	5	0	10	23	18
731	101	2	1	3	50.2	4	5	3	8	13	11
731	204	1	0	4	24	0	0	3	3	0	0
731	205	1	1	4	21.9	3	2	5	1	5	0
1097	101	2	0	3	43	2	3	2	1	0	6
1097	102	1	1	4	49.1	3	0	3	2	2	2
1097	203	1	0	4	23.5	1	4	1	3	2	2
1689	101	2	0	1	44.9	3	7	5	16	7	8
1689	102	1	2	3	47.8	1	8	24	22	14	8
1689	204	1	3	2	21.6	6	8	3	2	6	4
1906	101	2	4	1	67.8	27	23	29	39	19	16
1906	202	2	0	2	47.5	2	0	4	5	9	8
1906	203	1	1	2	50.2	12	8	8	11	9	13

Chapter 7
Longitudinal Models for Binary Data

In Chapter 6, we have discussed the stationary and nonstationary correlation models for count data, and estimated the effects of the covariates on the count responses, by taking the correlation structure into account. In this chapter, we deal with repeated binary responses. For example, there exists a longitudinal study on the health effects of air pollution, where wheezing status (1 = yes, 0 = no) of a large number of independent children are repeatedly recorded, along with maternal smoking status, family cleanliness status, level of chemicals used, and pet-owning status of the family. For $i = 1,\ldots,K$, and $t = 1,\ldots,T$, let y_{it} denote the binary response and $x_{it} = (x_{it1},\ldots,x_{itp})'$ denote the p-dimensional covariate vector collected at time point t from the ith individual. Similarly, one may be interested to study employment data for many individuals over a short period of $T = 4$ years. Here $y_{it} = 1$ may be used to indicate that the ith individual was unemployed at time point t, whereas $y_{it} = 0$ indicates that the individual was employed. In this example, x_{it}, the covariate vector, may consist of some of the important covariates such as gender, age, education level, geographic location, and marital status of the individual. Let β be the effect of x_{it} on y_{it}. Note that because $y_{i1},\ldots,y_{it},\ldots,y_{iT}$ are T repeated binary responses from the same individual, it is most likely that they are autocorrelated. The scientific concern is to find β, the effects of the covariates on the repeated binary responses, after taking their autocorrelations into account.

Note that there are also situations in practice, where the covariates of the ith individual may be time independent. We denote such covariates by $\tilde{x}_i = (x_{i1},\ldots,x_{ip})'$. This is a simpler special case of the general situation with time-dependent covariates x_{it}. Note that when the covariates are time dependent, the responses follow a nonstationary correlation model, whereas in the special case when covariates are time independent, the correlation model becomes stationary. In Section 7.1, we provide the marginal distributional properties of the binary response variable Y_{it} under the general situation when corresponding covariates are time dependent. In the same section, we discuss the estimation of β by pretending that the repeated binary responses are independent, even though in reality they are correlated. In Section 7.2, we discuss three selected binary correlation models, namely a multivariate density based (MBD) model due to Bahadur (1961), an autoregressive order 1 (AR(1)) type

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_7,

observation-driven dynamic (ODD) model [Kanter (1975)], and a linear dynamic conditional probability (LDCP) model [Qaqish (2003)]. These models are discussed first for the special case with time-independent covariates and then for the general case with time-dependent covariates. All of these three probability models produce the same marginal mean and the variance, and the MDB and LDCP models can accommodate any desired correlation structures, whereas the ODD model follows an AR(1) type structure. Note, however, that the ranges for correlations under all three models are restricted by probability conditions, the LDCP model being more flexible that accommodates correlations satisfying a wider range as compared to the other two models. In the same section, a numerical study is reported on the range performances of these probability model based correlations. In Section 7.3, we provide an autocorrelation class of correlation models for the stationary binary data. In the same section, we discuss the GQL inferences for the regression effects β, after taking the stationary correlation structure of the repeated data into account. In Section 7.4, we generalize the class of correlation structures to the nonstationary case. We consider a numerical example in Section 7.5 and illustrate the application of stationary correlation structure based model fitting to the nonstationary survey of labor and income dynamic (SLID) data collected by Statistics Canada. In Section 7.6, a stationary correlation structure based binary model is considered in a longitudinal clinical trial setup. The longitudinal adaptive design based weighted generalized quasi-likelihood (WGQL) inference is introduced for the estimation of the regression parameters including the treatment effects.

Note that the nonstationary binary models discussed in Section 7.4 accommodate specified marginal means and variances and a suitable class of nonstationary (i.e., time-dependent) correlation structures. In practice, there are, however, situations in the longitudinal setup, where the mean and the variance at a given time point may maintain some deterministic relationship with their past counterparts. To analyze this type of non-stationary longitudinal binary data, in Section 7.7, we discuss a nonlinear binary dynamic logit (BDL) model as opposed to the LDCP models from Section 7.2. This is quite interesting to point out that this NLDCP model for repeated binary data always accommodates correlations with full ranges from -1 to $+1$. In this BDL model setup, we consider several estimation approaches such as maximum likelihood (ML), GQL, and an optimal GQL (OGQL) approach for the estimation of the regression effects and a dynamic dependence parameter (an index for correlations), and study their properties through a simulation study. In the same section, longitudinal binary data on asthma status are analyzed by using the ML, GQL, and OGQL estimation approaches. In the same section, we demonstrate the application of a BDL model in a longitudinal adaptive clinical trial setup with possibly more than two treatments.

7.1 Marginal Model

Even though the T repeated binary responses for the ith individual are autocorrelated, marginal model based inferences either ignore the correlations and hence use the independence assumption, or use the 'working' correlations assumption (such as by the GEE approach) without modelling the correlations. The drawbacks of the GEE (generalized estimating equations) approach for the longitudinal count data analysis were discussed in Section 6.4.3, and we revisit this issue in Section 7.4 in brief for the longitudinal binary data analysis. For the purpose of using the 'working' independence assumption in estimating the regression effects β, we now write a standard logistic binary marginal density and provide some of its moment properties. The density can be used to obtain independence based ML estimate, whereas the moments are used to obtain MM (method of moments) and QL (quasi-likelihood) estimates.

For convenience we write the logistic binary distribution of Y_{it} in exponential density form given by

$$f(y_{it}) = \exp[\{y_{it}\theta_{it} - a(\theta_{it})\} + b^*(y_{it})], \tag{7.1}$$

which was also used for the Poisson case (see eqn. (6.1)), but unlike the Poisson case, we now have

$$a(\theta_{it}) = \log\{1+\exp(\theta_{it})\}, \text{ with } \theta_{it} = x'_{it}\beta. \tag{7.2}$$

Also in (7.1), $b^*(y_{it}) = 1$.

Let $a'(\theta_{it})$, $a''(\theta_{it})$, $a'''(\theta_{it})$, and $a''''(\theta_{it})$ be, respectively, the first−, second−, third− and the fourth-order derivatives of $a(\theta_{it})$ with respect to θ_{it}. By using the m.g.f. as for (6.2), or by direct calculations, one obtains the first four marginal moments of the binary variable as in the following lemma.

Lemma 7.1 The first four moments of the binary random variable Y_{it} under the exponential family density $(7.1)-(7.2)$ are given by

$$\begin{aligned}
\pi_{it} &= E[Y_{it}] = a'(\theta_{it}) = \frac{\exp(\theta_{it})}{1+\exp(\theta_{it})} \\
\sigma_{itt} &= \text{var}[Y_{it}] = a''(\theta_{it}) = \pi_{it}(1-\pi_{it}) \\
\tilde{\delta}_{itt} &= E[Y_{it}-\mu_{it}]^3 = a'''(\theta_{it}) = \pi_{it}(1-\pi_{it})(1-2\pi_{it}) \\
\tilde{\phi}_{itttt} &= E[Y_{it}-\mu_{it}]^4 = a''''(\theta_{it}) + 3\sigma_{itt}^2 = \pi_{it}(1-\pi_{it})\{1-3\pi_{it}(1-\pi_{it})\}.
\end{aligned} \tag{7.3}$$

We denote the marginal binary distribution given by $(7.1)-(7.2)$ as $Y_{it} \sim b(\pi_{it})$.

7.1.1 Marginal Model Based Estimation for Regression Effects

Note that as the Poisson and the binary densities are written in (6.1) and (7.1) in the same form of exponential family density, the MM, QL, and ML estimating equations for β under the binary model have the same expressions as those under the Poisson model discussed in Section 6.2. Thus, these equations are, respectively, given by

Method of Moments (MM)

$$\sum_{i=1}^{K}\sum_{t=1}^{T}[x_{it}(y_{it}-a'(\theta_{it}))]=0, \tag{7.4}$$

Quasilikelihood (QL) Method

$$\sum_{i=1}^{K}\sum_{t=1}^{T}[\frac{\partial a'(\theta_{it})}{\partial\beta}\frac{(y_{it}-a'(\theta_{it}))}{\mathrm{var}(Y_{it})}]=0, \tag{7.5}$$

and

Marginal Likelihood (ML) Method

$$\frac{\partial \log L}{\partial\beta}=\sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it}-a'(\theta_{it})]\frac{\partial\theta_{it}}{\partial\beta}=0, \tag{7.6}$$

where, under the binary model, by Lemma 7.1, we now have

$$a'(\theta_{it})=\pi_{it}=\frac{\exp(\theta_{it})}{1+\exp(\theta_{it})}\ \text{ and }\ \mathrm{var}(Y_{it})=\pi_{it}(1-\pi_{it}),\ \text{ with } \theta_{it}=x'_{it}\beta.$$

Note that all three approaches, namely MM (7.4), QL (7.5), and ML (7.6) estimating equations provide the same estimate for β, as they have the same estimating equation form given by

$$\sum_{i=1}^{K}[X'_i(y_i-\pi_i)]=0, \tag{7.7}$$

[see also (6.5)], where

$$y_i=(y_{i1},\ldots,y_{it},\ldots,y_{iT})',\quad X'_i=(x_{i1},\ldots,x_{it},\ldots,x_{iT}),$$

and

$$\pi_i=(\pi_{i1},\ldots,\pi_{it},\ldots,\pi_{iT})'\ \text{ with } \pi_{it}=\frac{\exp(x'_{it}\beta)}{1+\exp(x'_{it}\beta)}.$$

Let $\hat{\beta}$ be the solution of (7.7) for β. This estimate may be obtained by using the iterative equation

$$\hat{\beta}(r+1) = \hat{\beta}(r) + \left[\sum_{i=1}^{K} X_i' A_i X_i\right]_{(r)}^{-1} \left[\sum_{i=1}^{K} X_i'(y_i - \pi_i)\right]_{(r)}, \tag{7.8}$$

where

$$\begin{aligned} A_i &= \text{diag}[a''(\theta_{i1}), \ldots, a''(\theta_{it}), \ldots, a''(\theta_{iT})] \\ &= \text{diag}[\sigma_{i11}, \ldots, \sigma_{itt}, \ldots, \sigma_{iTT}] \\ &= \text{diag}[\pi_{i1}(1-\pi_{i1}), \ldots, \pi_{it}(1-\pi_{it}), \ldots, \pi_{iT}(1-\pi_{iT})], \end{aligned} \tag{7.9}$$

and $[\cdot]_{(r)}$ denotes that the expression within the brackets is evaluated at $\beta = \hat{\beta}(r)$, the rth iterative value for $\hat{\beta}$. Furthermore, similar to (6.7), it may be shown that $K^{1/2}(\hat{\beta} - \beta)$ is asymptotically multivariate Gaussian with zero mean vector and covariance matrix V_M^* given by

$$V_M^* = \text{limit}_{K\to\infty} K \left[\sum_{i=1}^{K} X_i' A_i X\right]^{-1} \left[\sum_{i=1}^{K} X_i' A_i^{1/2} C_i A_i^{1/2} X_i\right] \left[\sum_{i=1}^{K} X_i' A_i X\right]^{-1}, \tag{7.10}$$

where A_i is given by (7.9), and C_i is the true correlation matrix of y_i which may be unknown.

7.2 Some Selected Correlation Models for Longitudinal Binary Data

There is a long history on the modelling of correlated binary data in the time series setup. For example, one may refer to some of the early works such as by Bahadur (1961), Cox and Lewis (1966), Klotz (1973), Kanter (1975), Lindquist (1978), Keenan (1982), and Jacobs and Lewis (1983). Among the recent works, one may, for example, refer to the Markov dependence type linear dynamic conditional probability based model discussed by Qaqish (2003). Note that among these models, the multivariate binary density based model by Bahadur (1961) [see also Cox and Lewis (1966); Cox (1972); Prentice (1988), for example] and the Markov dependence based LDCP type model [e.g. Zeger, Liang and Self (1985)] are widely used. There are, however, two main difficulties with these models. First, even though the functional forms for the marginal means and variances remain the same for all time points, these models are developed such that they can accommodate any correlation structures. As opposed to the Gaussian type stationary correlation structure, nonstationary correlation structures are less familiar, therefore these models may be of very limited use in the nonstationary case. Secondly, even if the stationary correlations are used, the ranges for the correlations can be narrow, which, however, are needed to be satisfied for any inferences for the data. In the nonstationary case, the range restrictions pose more serious problems. In Section 7.2.1, we consider these

MBD and LDCP models both for the stationary and nonstationary cases. In Section 7.2.2, following Farrell and Sutradhar (2006), we compare the range performances for AR(1) type correlations under these two models, in the stationary case. Because there also exist an observation-driven dynamic AR(1) type model proposed by Kanter (1975), we consider this ODD model as well in the comparison of ranges for correlations. It is demonstrated that the LDCP model allows wider ranges for the correlations as opposed to the MBD and ODD models. In Section 7.3 and 7.5, we introduce a class of stationary autocorrelated models of low order which are similar to but different from the LDCP models.

7.2.1 *Bahadur Multivariate Binary Density (MBD) Based Model*

To develop a correlation model for the repeated binary responses $y_{i1},\ldots,y_{it},\ldots,Y_{iT}$ for all $i = 1,\ldots,K$, it is sufficient to develop such a model for a single individual. This is because the individuals are independent. We now, therefore, explain the MBD model for the ith individual only. The other two correlation models, namely, the ODD and LDCP models are also discussed for the ith individual.

7.2.1.1 Stationary Case

In the stationary setup, covariates are time independent. Recall that we use $x_{it} = \tilde{x}_i$ for all $t = 1,\ldots,T$, to represent such a time-independent covariate. This reduces the marginal probability π_{it} in Lemma 7.1 to $\tilde{\pi}_i$. To be more specific, in the stationary case

$$\tilde{\pi}_i = P(y_{it} = 1) = \frac{\exp(\tilde{x}_i'\beta)}{1+\exp(\tilde{x}_i'\beta)},$$

so that for $t = 1,\ldots,T$,

$$E[Y_{it}] = \tilde{\pi}_i, \text{ and } \operatorname{var}[Y_{it}] = \tilde{\pi}_i(1-\tilde{\pi}_i).$$

Next, suppose that for $u \neq t, u,t = 1,\ldots,T$, $c^*_{i,ut}$ denote a general pairwise correlation between y_{it} and y_{iu}. In the present stationary case, this correlation is independent of covariates $\tilde{x}_i$. For example, if the repeated observations follow a Gaussian type AR(1) stationary correlation model, then $c^*_{i,ut} \equiv c^*_{i,ut}(\rho) = \rho^{|t-u|}$ with ρ as the correlation index parameter. For $T \geq 2$, the MBD considered by Bahadur (1961) is written as

$$f(y_{i1},\ldots,y_{iT}) = \Pi_{t=1}^{T}\tilde{\pi}_i^{y_{it}}(1-\tilde{\pi}_i)^{1-y_{it}} \times \left[1+\sum_{t<u}^{T} c^*_{i,ut}\left(\frac{y_{iu}-\tilde{\pi}_i}{[\tilde{\pi}_i(1-\tilde{\pi}_i)]^{\frac{1}{2}}}\right)\left(\frac{y_{it}-\tilde{\pi}_i}{[\tilde{\pi}_i(1-\tilde{\pi}_i)]^{\frac{1}{2}}}\right)\right], \quad (7.11)$$

which, alternatively, can also be expressed in a simpler way as

$$f(y_{i1},\ldots,y_{iT}) = 1+[\sum_{t<u}^{T} c^*_{i,ut}(-1)^{y_{it}+y_{iu}}\tilde{\pi}_i^{2-y_{it}-y_{iu}}(1-\tilde{\pi}_i)^{y_{it}+y_{iu}}]/[\tilde{\pi}_i(1-\tilde{\pi}_i)], \quad (7.12)$$

where $y_{it} = 0,1$ for any i and all $t = 1,\ldots,T$.

Mean, Variance, and Covariance or Correlation Structures

It follows from the joint density (7.11) that the marginal mean and the variance of Y_{it} are $E[Y_{it}] = \tilde{\pi}_i$ and $\text{var}[Y_{it}] = \tilde{\pi}_i(1-\tilde{\pi}_i)$, respectively. Also, $c^*_{i,ut}$ is the correlation between y_{iu} and y_{it}. Note that to verify these functions, it is sufficient to check these properties from the bivariate density of y_{i1} and y_{i2}, for example. For this purpose, by summing over $y_{ij} = 0,1$, for $j = 3,\ldots,T$, one can write the bivariate density of y_{i1} and y_{i2}, from (7.11), as given by

$$f(y_{i1},y_{i2}) = \Pi_{t=1}^{2}\tilde{\pi}_i^{y_{it}}(1-\tilde{\pi}_i)^{1-y_{it}}\left[1+c^*_{i,12}\frac{(y_{i1}-\tilde{\pi}_i)(y_{i2}-\tilde{\pi}_i)}{\tilde{\pi}_i(1-\tilde{\pi}_i)}\right]. \quad (7.13)$$

This density provides

$$\begin{aligned} E[Y_{i1}] &= Pr[Y_{i1}=1] = f(y_{i1}=1,y_{i2}=0)+f(y_{i1}=1,y_{i2}=1) = \tilde{\pi}_i \\ E[Y_{i2}] &= Pr[Y_{i2}=1] = f(y_{i1}=0,y_{i2}=1)+f(y_{i1}=1,y_{i2}=1) = \tilde{\pi}_i, \end{aligned} \quad (7.14)$$

yielding the desired means. Next, because $E[Y_{it}^2] = E[Y_{it}]$ for binary y_{it}, it then follows that

$$\text{var}[Y_{it}] = \tilde{\pi}_i - \tilde{\pi}_i^2 = \tilde{\pi}_i(1-\tilde{\pi}_i).$$

Furthermore, it follows from the bivariate density (7.13) that

$$E[Y_{i1}Y_{i2}] = Pr[Y_{i1}=1,Y_{i2}=1] = f(y_{i1}=1,y_{i2}=1) = \tilde{\pi}_i^2 + c^*_{i,12}[\tilde{\pi}_i(1-\tilde{\pi}_i)], \quad (7.15)$$

yielding the desired correlation as

$$\begin{aligned} \text{corr}(Y_{i1},Y_{i2}) &= \frac{\text{cov}(Y_{i1},Y_{i2})}{[\text{var}[Y_{it}]\text{var}[Y_{it}]]^{\frac{1}{2}}} \\ &= c^*_{i,12}. \end{aligned} \quad (7.16)$$

Range for Correlation Index Parameter

Note that even if a specific form is considered for the pairwise correlations $c^*_{i,ut}$ in (7.11), the computation for the range of correlations gets complicated when T increases. For any T, the ranges are in general functions of marginal probabilities. For example, suppose that we consider the Gaussian AR(1) type correlation structure mentioned above, namely, $c^*_{i,ut} \equiv c^*_{i,ut}(\rho) = \rho^{|t-u|}$ for all individuals $i = 1,\ldots,K$, with ρ as a correlation index parameter. Note that this correlation index parameter ρ is also the lag 1 correlation. For $T = 2$, the correlation index parameter or the lag 1 correlation ρ has the range

$$\max\left[-\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}, -\frac{1-\tilde{\pi}_i}{\tilde{\pi}_i}\right] < \rho < 1, \tag{7.17}$$

which is quite different than the Gaussian type range -1 to $+1$. For $T > 2$, the range becomes more restricted, and it may be computed numerically such that a selected value of ρ satisfies the probability range restriction

$$0 < f(y_{i1}, \ldots, y_{it}, \ldots, y_{iT}) < 1, \tag{7.18}$$

for all $i = 1, \ldots, K$. Note, however, that this computation can be cumbersome. For example, to find the range for the lag 1 parameter ρ, under the AR(1) type correlation structure for T = 4 and a value of $\tilde{\pi}_i$ =0.40, (say), for all $i = 1, \ldots, K$, one needs to compute all 2^4 values of $f(\cdot)$ for each ρ = -0.999 (0.001) 0.999, and obtain the range under which all 2^4 values of $f(\cdot)$ are found to lie between 0 and 1. This range is $-0.262 < \rho < 0.449$; see Table 7.1 [see also Farrell and Sutradhar (2006)].

7.2.1.2 Nonstationary Case

Similar to (7.11), the Bahadur's MBD under the nonstationarity condition is written as

$$f(y_{i1}, \ldots, y_{iT}) = \Pi_{t=1}^{T} \pi_{it}^{y_{it}} (1-\pi_{it})^{1-y_{it}} \left[1 + \sum_{t<u}^{T} c_{i,ut}^{*} \frac{[(y_{iu}-\pi_{iu})(y_{it}-\pi_{it})]}{[\pi_{iu}(1-\pi_{iu})\pi_{it}(1-\pi_{it})]^{1/2}}\right], \tag{7.19}$$

where, for example, we may consider

$$\pi_{it} = \frac{\exp(x'_{it}\beta)}{1+\exp(x'_{it}\beta)}$$

with x_{it} as the time-dependent covariates for $t = 1, \ldots, T$, and for all $i = 1, \ldots, K$.

Note that one may exploit this MBD in (7.19) and show that

$$\begin{aligned} E[Y_{it}] = \pi_{it},\ \ \text{var}[Y_{it}] = \pi_{it}(1-\pi_{it})\ \text{ for } t = 1, \ldots, T \\ \text{corr}[Y_{iu}, Y_{it}] = c_{i,ut}^{*},\ \text{ for } u < t; t = 2, \ldots, T. \end{aligned} \tag{7.20}$$

Further note that the correlations $c_{i,ut}^{*}$ in (7.19) and (7.20) are nonstationary and in general they may be denoted by $c_{i,ut}^{(ns)}(x_{iu}, x_{it}, \rho)$ (see (6.46) for the Poisson data), ρ being a correlation index parameter. However, even if a stationary correlation structure such as $c_{i,ut}^{*} \equiv c_{i,ut}^{*}(\rho) = \rho^{|t=u|}$ is used in place of nonstationary correlations, finding the range for ρ by exploiting (7.19) is naturally much more difficult than finding its range by exploiting the stationary density (7.11) [see also (7.18)]. This is because, unlike in the stationary case, the range for ρ will also depend on the values of x_{it}. For example, for $T = 2$, ρ has to satisfy the range restriction

$$\max\left[-\left\{\frac{\pi_{i1}\pi_{i2}}{\pi^c_{i1}\pi^c_{i2}}\right\}^{1/2}, -\left\{\frac{\pi^c_{i1}\pi^c_{i2}}{\pi_{i1}\pi_{i2}}\right\}^{1/2}\right] < \rho < \min\left[\left\{\frac{\pi_{i1}\pi^c_{i2}}{\pi_{i2}\pi^c_{i1}}\right\}^{1/2}, \left\{\frac{\pi_{i2}\pi^c_{i1}}{\pi_{i1}\pi^c_{i2}}\right\}^{1/2}\right], \tag{7.21}$$

where $\pi^c_{it} = 1 - \pi_{it}$ for $t = 1, 2$. For $T > 2$, finding the range restriction for ρ is much more complicated. This is one of the reasons why we do not follow the MBD any more in the chapter for the inference purpose, that is, for the estimation of the regression parameter β.

7.2.2 *Kanter Observation-Driven Dynamic (ODD) Model*

7.2.2.1 Stationary Case

Kanter (1975) has introduced an observation-driven dynamic correlated binary model for stationary time series data. In the context of the present longitudinal setup, suppose that y_{i1} is binary with probability $\tilde{\pi}_i$. Further suppose that for $t = 2, \ldots, T$, s_{it} is a binary random variable with

$$Pr(s_{it} = 1) = \gamma_1, \text{ with } 0 < \gamma_1 < 1,$$

and d_{it} is another binary random variable with

$$Pr(d_{it} = 1) = \xi^*_i = \tilde{\pi}_i(1 - \gamma_1)/(1 - 2\gamma_1\tilde{\pi}_i).$$

Following Kanter (1975), one may then generate the AR(1)-type correlated responses $y_{i1}, \ldots, y_{iT}$ by using the model

$$y_{it} = s_{it}\{y_{i,t-1} \oplus d_{it}\} + (1 - s_{it})d_{it}, \text{ for } t = 2, \ldots, T, \tag{7.22}$$

where $\oplus$ denotes addition modulo 2. Now, if $y_{i,t-1}$, s_{it}, and d_{it} are assumed to be independent, it follows from (7.22) that $y_{it}(t = 2, \ldots, T)$ has a binary distribution with $Pr(y_{it} = 1) = \tilde{\pi}_i$, which is the same distribution as that of y_{i1}.

Mean, Variance, and Covariance or Correlation Structures

Because $y_{i1} \sim b(\tilde{\pi}_i)$ and also $y_{it} \sim b(\tilde{\pi}_i)$ for $t = 2, \ldots, T$, by (7.22), it then follows that

$$\begin{aligned} E[Y_{it}] &= \pi_{it} = \tilde{\pi}_i, \text{ for } t = 1, \ldots, T \\ \text{var}[Y_{it}] &= \tilde{\pi}_i(1 - \tilde{\pi}_i) \text{ for } t = 1, \ldots, T. \end{aligned} \tag{7.23}$$

It can further be shown by (7.22) that for

$$\rho_i = \gamma_1(1 - 2\tilde{\pi}_i)/(1 - 2\gamma_1\tilde{\pi}_i), \tag{7.24}$$

the correlation between y_{it} and y_{iu} is given by

$$Corr(Y_{it}, Y_{iu}) = c^*_{i,ut}(\rho) = \rho_i^{|t-u|}, \text{ for } t \neq u,$$

which appears to be similar to the lag $|t-u|$ autocorrelation under a Gaussian autoregressive process. Note that here ρ_i is the lag 1 correlation among the binary responses of the ith individual, whereas γ_1 that defines ρ_i as in (7.24) is referred to as the correlation index parameter.

Range for Correlation Index Parameter:

Note that $Pr(d_{it} = 1) = \xi_i^*$ has the range $0 < \xi_i^* < 1$, implying that the correlation index parameter γ_1 must satisfy the range restriction

$$0 < \gamma_1 < \min\{\frac{1-\tilde{\pi}_i}{\tilde{\pi}_i}, 1\}, \text{ for } i = 1, \ldots, K. \tag{7.25}$$

Further note that this range restriction must be satisfied when computing the estimate of β, involved in all $\tilde{\pi}_i$ for $i = 1, \ldots, K$, by using ρ_i as as function of γ_1.

Range for Lag 1 Correlation

Now to find the range restrictions for the lag 1 correlation ρ_i as a function of $\tilde{\pi}_i$, we substitute the value for $\tilde{\pi}_i$ into the formula for ξ_i^*. We then compute ranges for γ_1 by using $0 < \xi_i^* < 1$ for all $i = 1, \ldots, K$. These ranges for γ_1 are then used in (7.24) to obtain the ranges for ρ_i under the AR(1) process, as follows:

$$\begin{aligned} \rho_i = 0, &\text{ for } \tilde{\pi}_i = 0.5 \\ 0 < \rho_i < 1, &\text{ for } 0 < \tilde{\pi}_i < 0.5 \\ -\min_i\left[\frac{1-\tilde{\pi}_i}{\tilde{\pi}_i}\right] \frac{1-2\tilde{\pi}_i}{2\tilde{\pi}_i \min_i\left[\frac{1-\tilde{\pi}_i}{\tilde{\pi}_i}\right] - 1} < \rho_i < 0, &\text{ for } 0.5 < \tilde{\pi}_i < 1 \end{aligned} \tag{7.26}$$

Note that because γ_1 is considered to be a common parameter for the correlation structures for all K individuals, the range of this correlation index parameter as shown in (7.25) depends on the values of $\tilde{\pi}_i$ for all $i = 1, \ldots, K$. Consequently, as given in (7.26), the range for ρ_i, the lag 1 correlation for the ith individual also depends on the values of $\tilde{\pi}_i$ for all $i = 1, \ldots, K$.

Further note that in the context of stationary binary time series generated by (7.22), one deals with $K = 1$. For such a case with $i = 1$ only, by writing

$$\tilde{\pi} \equiv \tilde{\pi}_1, \quad \rho \equiv \rho_1,$$

one may simplify (7.26) and obtain the range for the lag 1 correlation ρ as

$$\begin{aligned} \rho = 0, &\text{ for } \tilde{\pi} = 0.5 \\ 0 < \rho < 1, &\text{ for } 0 < \tilde{\pi} < 0.5 \end{aligned}$$

$$-\frac{1-\tilde{\pi}}{\tilde{\pi}} < \rho < 0, \text{ for } 0.5 < \tilde{\pi} < 1. \tag{7.27}$$

Alternatively, the range restrictions in (7.27) are also valid in a specialized longitudinal setup with

$$\tilde{\pi}_i = \tilde{\pi}, \text{ for all } i = 1, \ldots, K.$$

This is because in such a case one naturally uses

$$\rho = \gamma_1(1-2\tilde{\pi})/(1-2\gamma_1\tilde{\pi}), \tag{7.28}$$

as the special case of (7.24). Note, however, that this special case with the same stationary binary probability for the responses of all K individuals is most unlikely in practice.

7.2.2.2 Non-stationary Case

In the nonstationary case, the dynamic relationship (7.22) still holds with the same probability function for the s_{it} binary variable; that is, $Pr(s_{it} = 1) = \gamma_1$, with $0 < \gamma_1 < 1$, but d_{it} is now a binary random variable with

$$Pr(d_{it} = 1) = \xi^*_{it} = \pi_{it}(1-\gamma_1)/(1-2\gamma_1\pi_{it}), \tag{7.29}$$

where

$$\pi_{it} = \frac{\exp(x'_{it}\beta)}{1+\exp(x'_{it}\beta)}.$$

The nonstationary mean and the variance of y_{it} are given by

$$E[Y_{it}] = \pi_{it}, \text{ and } \mathrm{var}[Y_{it}] = \pi_{it}(1-\pi_{it}),$$

respectively. Next, for the purpose of computing the lag $w = 1, 2, \ldots, T-1$, correlations, one may compute the lag w autocovariance by using the formula

$$\mathrm{cov}[Y_{i,t-w}, Y_{it}] = Pr[Y_{i,t-w} = 1,\ Y_{it} = 1] + \pi_{i,t-w}\pi_{it}, \tag{7.30}$$

for $t = w+1, \ldots, T$. For $w = 1$, and 2, for example, the joint probabilities are given by

$$Pr(Y_{i,t-1} = 1,\ Y_{it} = 1) = Pr(Y_{i,t-1} = 1)Pr(Y_{it} = 1|Y_{i,t-1} = 1), \tag{7.31}$$

and

$$\begin{aligned} Pr(Y_{i,t-2} = 1,\ Y_{it} = 1) = Pr(Y_{i,t-2} = 1) & \\ \times [Pr(Y_{i,t-1} = 1|Y_{i,t-2} = 1)Pr(Y_{it} = 1|Y_{i,t-1} = 1) & \\ + Pr(Y_{i,t-1} = 0|Y_{i,t-2} = 1) & \\ \times Pr(Y_{it} = 1|Y_{i,t-1} = 0)], & \end{aligned} \tag{7.32}$$

respectively, where $Pr[Y_{i,t-w} = 1] = \pi_{i,t-w}$, and the conditional probabilities have the formulas

$$\begin{aligned}
Pr(Y_{i,t-w} = 1|Y_{i,t-w-1} = 1) &= \gamma_1 + (1-2\gamma_1)\xi_{i,t-w} \\
Pr(Y_{i,t-w} = 0|Y_{i,t-w-1} = 1) &= 1-\gamma_1 - (1-2\gamma_1)\xi_{i,t-w} \\
Pr(Y_{i,t-w} = 1|Y_{i,t-w-1} = 0) &= \xi_{i,t-w} \\
Pr(Y_{i,t-w} = 0|Y_{i,t-w-1} = 0) &= 1-\xi_{i,t-w}
\end{aligned} \tag{7.33}$$

with

$$\xi_{i,t-w} = \frac{\pi_{i,t-w} - \gamma_1 \pi_{i,t-w-1}}{1-2\gamma_1 \pi_{i,t-w-1}}. \tag{7.34}$$

Note that because the probability ξ^*_{it} in (7.29) must lie between 0 and 1, it then follows that in the non-stationary case, the correlation index parameter γ_1 must satisfy the range restriction

$$0 < \gamma_1 < \min\left[\frac{\pi_{it}}{\pi_{i,t-1}}, \frac{1-\pi_{it}}{\pi_{i,t-1}}, \frac{1}{2\pi_{i,t-1}}\right]$$

for all $i = 1, \ldots, K$; $t = 2, \ldots, T$.

7.2.3 *A Linear Dynamic Conditional Probability (LDCP) Model*

7.2.3.1 Stationary Case

To model the correlated binary data, many authors [see Zeger, Liang, and Self (1985), e.g.] have used the Markovian or AR(1) type LDCP model given by

$$\begin{aligned}
Y_{i1} &\sim b(\tilde{\pi}_i) \\
Pr[Y_{it} &= 1|Y_{i,t-1} = y_{i,t-1}] = \tilde{\pi}_i + \rho(y_{i,t-1} - \tilde{\pi}_i), \text{ for } t = 2, \ldots, T.
\end{aligned} \tag{7.35}$$

This model produces the marginal mean and the variance of y_{it} for all $t = 1, \ldots, T$, as

$$\begin{aligned}
E[Y_{it}] &= \tilde{\pi}_i \\
\text{var}[Y_{it}] &= \tilde{\pi}_i(1-\tilde{\pi}_i),
\end{aligned} \tag{7.36}$$

and for $u < t$, the Gaussian type lag $t-u$ autocorrelation as

$$\text{Corr}[Y_{iu}, Y_{it}] = c^*_{i,ut}(\rho) = \rho^{t-u}, \tag{7.37}$$

but, unlike the Gaussian case, lag 1 correlation ρ must satisfy the range restriction

$$\max_i\left[-\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}, -\frac{1-\tilde{\pi}_i}{\tilde{\pi}_i}\right] \le \rho \le 1. \tag{7.38}$$

A General LDCP Model

There exists a generalization to the LDCP model (7.35) to accommodate any specified stationary correlation structure, along with the specified mean and variance. One may refer to Qaqish (2003), for example, for this generalization, where for the stationary case, the general model is written as

$$Pr(y_{it} = 1|y_{i,t-1},\ldots,y_{i1}) = \tilde{\pi}_i + \sum_{j=1}^{t-1} b_{i,tj}(y_{ij} - \tilde{\pi}_i), \text{ for } t = 2,\ldots,T, \tag{7.39}$$

with $b_{i,tj}$ as the dependence parameters that must satisfy the range restriction $0 < Pr(y_{it} = 1|y_{i,t-1},\ldots,y_{i1}) < 1$. It is clear from the model (7.39) that the marginal mean and the variance of Y_{it} are given by

$$E[Y_{it}] = \tilde{\pi}_i, \text{ var}[Y_{it}] = \tilde{\pi}_i(1-\tilde{\pi}_i).$$

Furthermore, similar to the Bahadur MBD model, this model (7.39) allows any specified correlation structure. Let $b^*_{i,t-1} = (b_{i,t1},\ldots,b_{i,t,t-1})'$ be the $t-1$-dimensional vector of dependence parameters. Also, consider

$$\begin{aligned} y^*_{i,t-1} &= [y_{i1},\ldots y_{i,t-1}]' : (t-1)\times 1 \\ A^*_i &= \text{diag}[a^*_{i,11},\ldots,a^*_{i,t-1,t-1}], \text{ with } a^*_{i,jj} = \text{var}(y_{ij}) = \tilde{\pi}_i(1-\tilde{\pi}_i), \end{aligned} \tag{7.40}$$

leading to the covariance matrix of $y^*_{i,t-1}$ as

$$\text{var}(Y^*_{i,t-1}) = (A^*_i)^{1/2} C^*_i (A^*_i)^{1/2}, \tag{7.41}$$

where

$$C^*_i = (c^*_{i,jk})$$

is the correlation matrix, with $c^*_{i,jk}$ being the (j,k)th $(j,k = 1,\ldots,t-1)$ element. One further obtains

$$\text{cov}(Y^*_{i,t-1}, Y_{it}) = (\sqrt{a^*_{i,tt}a^*_{i,11}}c^*_{i,1t},\ldots,\sqrt{a^*_{i,tt}a^*_{i,t-1,t-1}}c^*_{i,t-1,t})'. \tag{7.42}$$

Now by combining (7.41) and (7.42), one may compute the dependence vector $b^*_{i,t-1}$ as the function of correlations by using

$$b^*_{i,t-1} = [\text{var}(Y^*_{i,t-1})]^{-1}\text{cov}(Y^*_{i,t-1}, Y_{it}). \tag{7.43}$$

For example, one may compute this vector by using the Gaussian type stationary AR(1) correlation structure, namely

$$(c^*_{i,jk})(\rho) = \rho^{|j-k|}, \quad \text{for all } j \neq k,$$

ρ being the correlation index or lag 1 correlation parameter. Note, however, that because the model (7.39) has to satisfy $0 < Pr(y_{it} = 1|y_{i,t-1},\ldots,y_{i1}) < 1$, the AR(1) lag 1 correlation parameter ρ must satisfy the range restriction

$$\max_i \left[-\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}, -\frac{1-\tilde{\pi}_i}{\tilde{\pi}_i} \right] < \rho < 1.$$

Further note that the range restrictions for the correlations under other possible stationary Gaussian processes such as Gaussian MA(1) and exchangeable (equicorrelations), may be found similarly. See, for example, Qaqish (2003, Section 4).

7.2.3.2 Nonstationary Case

In the nonstationary case, that is, when the covariates are time dependent, following (7.39), one writes the LDCP model as

$$Pr(y_{it} = 1 | y_{i,t-1}, \ldots, y_{i1}) = \pi_{it} + \sum_{j=1}^{t-1} b_{i,tj}(y_{ij} - \pi_{it}), \text{ for } t = 2, \ldots, T, \tag{7.44}$$

where $\pi_{it} = \exp(x'_{it}\beta)/[1 + \exp(x'_{it}\beta)]$. As far as the dependence parameters are concerned, the dependence parameters vector $b^*_{i,t-1}$ may still be computed by using the formula in (7.43) given for the stationary case, except that unlike in the stationary case, one now uses

$$a^*_{i,jj} = \text{var}(Y_{ij}) = \pi_{ij}(1 - \pi_{ij}), \text{ for all } j = 1, \ldots, t-1.$$

Note that even if one uses the stationary correlation structure such as

$$(c^*_{i,jk})(\rho) = \rho^{|j-k|}, \quad \text{for all } j \neq k$$

in the nonstationary model (7.44) for the computation of the dependence parameter vector, the range restriction for the correlation index parameter ρ, or equivalently finding the range restrictions for the dynamic dependence parameters $b_{i,tj}$ would be much more complicated than in the stationary case. We do not pursue this complicated case any more. One may be referred to Qaqish (2003), for example, for the range restrictions under the nonstationary models with stationary correlation structures. Further note that in Section 7.3 we discuss simpler nonstationary binary models with both stationary and nonstationary correlation structures.

7.2.4 A Numerical Comparison of Range Restrictions for Correlation Index Parameter Under Stationary Binary Models

In the notation of Sections 7.2.1 – 7.2.3, Qaqish (2003) as well as Farrell and Sutradhar (2006) have considered the correlation structures for one individual, i.e., for

$K = 1$. For simplicity, we also consider this special case here and by writing $\tilde{\pi} = \tilde{\pi}_i$, for $i = 1$, we compute the range restrictions for the correlation index parameter ρ under a stationary AR(1) process, where the correlation structure is defined as

$$Corr(y_{it}, y_{iu}) = c^*_{i,ut}(\rho) = \rho^{|t-u|}.$$

To be specific, we compute these ranges by exploiting (7.18), (7.27), and (7.38) under the MBD, ODD, and LDCP models, respectively. These ranges for $T = 4$ and

$$\tilde{\pi} = 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ \text{and}\ 0.9$$

from Farrell and Sutradhar (2006, Table 1) are displayed in Table 7.1.

Table 7.1 For different values of $\tilde{\pi}$, range restrictions on ρ for each of the MDB (Bahadur), ODD (Kanter), and LDCP (Qaqish) models that are based on a stationary process with T = 4 and an AR(1) correlation structure.

$\tilde{\pi}$	MDB	ODD	LDCP
0.40	(−0.262, 0.449)	(0.000, 1.000)	(−0.667, 1.000)
0.50	(−0.430, 0.430)	(0.000, 0.000)	(−1.000, 1.000)
0.60	(−0.262, 0.449)	(-0.667, 0.000)	(−0.667, 1.000)
0.70	(−0.158, 0.503)	(-0.429, 0.000)	(−0.429, 1.000)
0.80	(−0.088, 0.486)	(-0.250, 0.000)	(−0.250, 1.000)
0.90	(−0.038, 0.442)	(-0.111, 0.000)	(−0.111, 1.000)

The results illustrate that the LDCP model provides the widest acceptable range for the correlation parameter at all values of $\tilde{\pi}$. In particular, the range restrictions for the MDB and ODD models are always more confining and entirely contained within the corresponding range for the LDCP model, regardless of the value of $\tilde{\pi}$. However, although the LDCP model handles all values of positive correlation regardless of the value of $\tilde{\pi}$, the only instance where the restriction for ρ under the LDCP model is over the entire range from -1 to 1 is when $\tilde{\pi} = 0.5$. In fact, the LDCP model restrictions allow for less of the range of negative correlation as $\tilde{\pi}$ moves farther away from 0.5, to the point where only negative correlations close to zero are permitted as $\tilde{\pi}$ approaches zero or one. Note that it was further indicated by Farrell and Sutradhar (2006) that the relative performance of the three models for the stationary equicorrelations and MA(1) models is similar to that for the AR(1) structure, but the ranges for the correlation parameters are more restrictive.

Because of the fact that among the three models, the LDCP model provides wider ranges for the correlation index parameter, and also because in practice, one frequently encounters low-order correlation structures, in Section 7.3, we consider such low-order correlation models, namely AR(1), MA(1), and equicorrelation LDCP models and deal with inferences for the stationary case. These stationary models are next generalized to the nonstationary models in Section 7.4. The inferences for the nonstationary case are also given in the same section.

7.3 Low-Order Autocorrelation Models for Stationary Binary Data

Similar to the stationary autocorrelation models for count data discussed in the last chapter (see Section 6.3), in this section, we consider three low-order stationary autocorrelation models, namely AR(1), MA(1), and EQC (equicorrelations) models for repeated binary data. All three models are founded on the same idea of writing the conditional probability for the current binary response in a linear dynamic form similar to that of (7.39). Note, however, that among these three models, we write the AR(1) model as a direct special case of (7.39), whereas the other two models are developed in a similar but different fashion. In fact, in the later two models, the conditional probability is expressed in a time series concept based MA(1) and EQC linear forms. For details on three models both in stationary and nonstationary setup, we refer to Sutradhar (2010). We now first describe the stationary models as follows. The nonstationary models are discussed in Section 7.4.

7.3.1 Binary AR(1) Model

This model and its basic marginal properties such as the mean and the variance, and its joint or product moment properties, namely autocorrelations, are described through equations from (7.35) to (7.38).

7.3.2 Binary MA(1) Model

Suppose that the repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are generated following the probability relationship

$$\begin{aligned} Pr[Y_{i1} = 1] &= \tilde{\pi}_i \\ Pr[Y_{it} = 1 | d_{it}, d_{i,t-1}] &= d_{it} + \rho d_{i,t-1}, \text{ for } t = 2, \ldots, T, \end{aligned} \tag{7.45}$$

where d_{it}s are independently distributed with mean ξ_i^* and variance η_i; that is,

$$d_{it} \overset{id}{\sim} \left[\xi_i^* = \frac{\tilde{\pi}_i}{1+\rho}, \ \eta_i = \tilde{\pi}_i(1-\tilde{\pi}_i) \right], \tag{7.46}$$

for all $t = 1, \ldots, T$. Note that the binary MA(1) model (7.45) is similar but different than the well-known Gaussian MA(1) model. This is because, in the Gaussian case, one uses the observation driven model, i.e., $y_{it} = d_{it} + \rho d_{i,t-1}$, whereas (7.45) is a conditional probability model. The distributional assumptions in two cases are also similar but different.

As far as the marginal properties of the model (7.45) are concerned, one obtains the means as

$$\begin{aligned} E[Y_{i1}] &= \mu_{i1} \\ E[Y_{it}] &= E_{d_{it},d_{i,t-1}} E[Y_{it}|d_{it},d_{i,t-1}] \\ &= E_{d_{it},d_{i,t-1}}[d_{it} + \rho d_{i,t-1}] \\ &= \frac{\tilde{\pi}_i}{1+\rho} + \rho\frac{\tilde{\pi}_i}{1+\rho} \\ &= \tilde{\pi}_i, \;\; \text{for}\; t = 2,\ldots,T, \end{aligned} \tag{7.47}$$

and the variances as

$$\begin{aligned} \text{var}[Y_{i1}] &= \tilde{\pi}_i[1-\tilde{\pi}_i] \\ \text{var}[Y_{it}] &= E_{d_{it},d_{i,t-1}}\text{var}[Y_{it}|d_{it},d_{i,t-1}] + \text{var}_{d_{it},d_{i,t-1}} E[Y_{it}|d_{it},d_{i,t-1}] \\ &= E_{d_{it},d_{i,t-1}}\left[E(Y_{it}^2|d_{it},d_{i,t-1}) - \{E(Y_{it}|d_{it},d_{i,t-1})\}^2\right] \\ &\quad + \text{var}_{d_{it},d_{i,t-1}}[d_{it}+\rho d_{i,t-1}] \\ &= E_{d_{it},d_{i,t-1}}[\{d_{it}+\rho d_{i,t-1}\} - \{d_{it}+\rho d_{i,t-1}\}^2] \\ &\quad + \text{var}_{d_{it},d_{i,t-1}}[d_{it}+\rho d_{i,t-1}] \\ &= E_{d_{it},d_{i,t-1}}[d_{it}+\rho d_{i,t-1}] - [E_{d_{it},d_{i,t-1}}\{d_{it}+\rho d_{i,t-1}\}]^2 \\ &= \tilde{\pi}_i - \tilde{\pi}_i^2 = \tilde{\pi}_i(1-\tilde{\pi}_i), \;\; \text{for}\; t=2,\ldots,T, \end{aligned} \tag{7.48}$$

respectively.

Next, for $u < t$, by using the model relationship (7.45), one may write

$$\begin{aligned} \text{cov}(Y_{iu},Y_{it}) &= E_{d_{iu},d_{i,u-1},d_{it},d_{i,t-1}}\text{cov}[(Y_{iu},Y_{it})|d_{iu},d_{i,u-1},d_{it},d_{i,t-1}] \\ &\quad + \text{cov}_{d_{iu},d_{i,u-1},d_{it},d_{i,t-1}}[d_{iu}+\rho d_{i,u-1}, d_{it}+\rho d_{i,t-1}] \\ &= \text{cov}_{d_{iu},d_{i,u-1},d_{it},d_{i,t-1}}[d_{iu}+\rho d_{i,u-1}, d_{it}+\rho d_{i,t-1}]. \end{aligned} \tag{7.49}$$

It then follows from (7.46) and (7.48) – (7.49) that

$$\text{corr}(Y_{iu},Y_{it}) = c_{i,ut}^*(\rho) = \begin{cases} \frac{\rho\,\text{var}(d_{iu})}{\sqrt{\text{var}(d_{iu})}\sqrt{\text{var}(d_{it})}} & \text{for}\; t-u=1 \\ 0 & \text{for}\; (t-u)>1 \end{cases} \tag{7.50}$$

$$= \begin{cases} \rho & \text{for}\; |u-t| = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{7.51}$$

Range for Lag 1 Correlation

The range for the lag 1 correlation or correlation index parameter depends on the nature of the model. The stationary MA(1) model given by (7.45) – (7.46) is semi-parametric by nature. This is because no distributional assumption is made in (7.46) for the d_{it} variable. Thus it becomes impossible to find the range. Now by making a reasonable distributional assumption (similar to the Gaussian type model) such as

$$d_{it} \overset{iid}{\sim} N\left[\xi_i^* = \frac{\tilde{\pi}_i}{1+\rho},\ \eta_i = \tilde{\pi}_i(1-\tilde{\pi}_i)\right],$$

one may attempt to find the range for ρ, which is, however, also not easy. Note, however, that in practice, when a consistent estimation method is used to compute the lag correlations such as in (7.51), the range issue may not be a big problem. Nevertheless, it is desirable to compute the range for the validity of the correlation interpretation as well as for the estimation of the main regression parameter β.

Note that as opposed to the probabilistic range, that is, the range for ρ satisfying the probability limits 0 to 1 for the conditional probability in (7.45), there exist procedures to find the weak stationary range for ρ by using the condition that the correlation matrix

$$C_i^*(\rho) = \begin{bmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{7.52}$$

defined by (7.51) be positive definite. A similar procedure is used by Qaqish (2003) to find the range for ρ, where the general dynamic model (7.39) is fitted with a specified MA(1) correlation structure. To be specific, suppose that the linear dynamic model (7.39) is used to define the binary MA(1) process with correlation structure (7.52), ρ being the lag 1 correlation. It can be shown that [see Qaqish (2003. eqn. (7)]

$$b_{i,tj} = \left[\frac{\sigma_{itt}}{\sigma_{ijj}}\right]^{\frac{1}{2}} \frac{a^j - a^{-j}}{a^{-t} - a^t} \text{ for } j = 1,\ldots,t-1;\ i = 1,\ldots,K, \tag{7.53}$$

where $\sigma_{ijj} = \tilde{\pi}_i(1-\tilde{\pi}_i)$ for all $j = 1,\ldots,T$, and $a \doteq \{(1-4\rho^2)^{1/2} - 1\}/(2\rho)$. Then, for a given i, the range for ρ can be computed by satisfying

$$L_T \le \tilde{\pi}_i \le 1 - L_T, \tag{7.54}$$

where

$$L_T = \begin{cases} \frac{a-a^T}{1-a^{T+1}} & (\rho < 0) \\ \frac{a^2-a^T}{(1+a)(1-a^{T+1})} & (\rho > 0,\ T \text{ even}) \\ \frac{(a-a^T)^2}{(1+a)(1-a^{T+1})(1-a^T)} & (\rho > 0,\ T \text{ odd}). \end{cases} \tag{7.55}$$

Suppose that (7.54)-(7.55) leads to the range for ρ as

$$c_{l,i} < \rho < c_{u,i},$$

for the ith $(i = 1,\ldots,K)$ individual. It then follows that the ρ will have the range

$$\max_i[c_{l,i}] < \rho < \min_i[c_{u,i}]. \tag{7.56}$$

For $i = 1$, the ranges for ρ under the model (7.39) are computed by Qaqish (2003, Table 1) [see also Farrell and Sutradhar (2006)] for different possible values of $\tilde{\pi}_1$.

7.3.3 Binary Equicorrelation (EQC) Model

Suppose that y_{i0} is an unobservable initial binary response with its mean $\tilde{\pi}_i$. Also, suppose that all repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are generated following the probability relationship

$$Pr[Y_{it} = 1|y_{i0}] = \tilde{\pi}_i + \rho(y_{i0} - \tilde{\pi}_i), \text{ for } t = 1, \ldots, T. \tag{7.57}$$

It then follows that

$$\begin{aligned} E[Y_{it}] &= E_{Y_{i0}} E[Y_{it}|Y_{i0}] = E_{Y_{i0}}[Pr(Y_{it} = 1|Y_{i0})] \\ &= E_{Y_{i0}}[\tilde{\pi}_i + \rho(Y_{i0} - \tilde{\pi}_i)] = \tilde{\pi}_i, \end{aligned} \tag{7.58}$$

and

$$\begin{aligned} \text{var}[Y_{it}] &= E_{Y_{i0}}[\text{var}\{Y_{it}|Y_{i0}\}] + \text{var}_{Y_{i0}}[E\{Y_{it}|Y_{i0}\}] \\ &= E_{Y_{i0}}[\{\tilde{\pi}_i + \rho(Y_{i0} - \tilde{\pi}_i)\}(1 - \{\tilde{\pi}_i + \rho(Y_{i0} - \tilde{\pi}_i)\})] + \text{var}_{Y_{i0}}[\tilde{\pi}_i + \rho(Y_{i0} - \tilde{\pi}_i)] \\ &= E_{Y_{i0}}\{\tilde{\pi}_i + \rho(Y_{i0} - \tilde{\pi}_i)\} - [E_{Y_{i0}}\{\tilde{\pi}_i + \rho(Y_{i0} - \tilde{\pi}_i)\}]^2 \\ &= \tilde{\pi}_i(1 - \tilde{\pi}_i). \end{aligned} \tag{7.59}$$

These means in (7.58) and the variances in (7.59) under the EQC model are the same as those of the AR(1) binary process given in (7.36), and also of the MA(1) binary process given in (7.47) – (7.48).

For $u \neq t$, by using the model relationship (7.57), one may write

$$\begin{aligned} \text{cov}(Y_{iu}, Y_{it}) &= E_{Y_{i0}} \text{cov}[(Y_{iu}, Y_{it})|Y_{i0}] \\ &\quad + \text{cov}_{Y_{i0}}[(\tilde{\pi}_i + \rho(Y_{i0} - \tilde{\pi}_i)), (\tilde{\pi}_i + \rho(Y_{i0} - \tilde{\pi}_i))] \\ &= \text{cov}_{Y_{i0}}[(\rho(Y_{i0} - \tilde{\pi}_i)), (\rho(Y_{i0} - \tilde{\pi}_i))] \\ &= \left[\rho^2 \tilde{\pi}_i(1 - \tilde{\pi}_i)\right] \\ &\equiv \left[\rho^2 \tilde{\pi}_i(1 - \tilde{\pi}_i)\right], \end{aligned} \tag{7.60}$$

yielding the correlations as

$$\begin{aligned} \text{corr}(Y_{iu}, Y_{it}) = c^*_{i,ut}(\rho) &= \frac{\left[\rho^2 \tilde{\pi}_i(1 - \tilde{\pi}_i)\right]}{\left[\sqrt{\tilde{\pi}_i(1 - \tilde{\pi}_i)}\sqrt{\tilde{\pi}_i(1 - \tilde{\pi}_i)}\right]} \\ &= \rho^2. \end{aligned} \tag{7.61}$$

Table 7.2 A class of stationary correlation models for longitudinal binary data and basic properties.

Model	Dynamic Relationship	Mean, Variance & Correlations
AR(1)	$Pr[Y_{i1}=1]=\tilde{\pi}_i$ $Pr[Y_{it}=1\|y_{i,t-1}]=\tilde{\pi}_i+\rho[y_{i,t-1}-\tilde{\pi}_i], t=2,\ldots$	$E[Y_{it}]=\tilde{\pi}_i$ $\text{var}[Y_{it}]=\tilde{\pi}_i(1-\tilde{\pi}_i)$ $\text{corr}[Y_{it},Y_{i,t+\ell}]=\rho_\ell$ $=\rho^\ell$
MA(1)	$Pr[Y_{it}=1\|d_{it},d_{i,t-1}]=d_{it}+\rho d_{i,t-1}, t=1,\ldots$ $d_{it}\sim[\text{mean}=\tilde{\pi}_i/(1+\rho), \text{var}=\tilde{\pi}_i(1-\tilde{\pi}_i)], t=0,1,\ldots$	$E[Y_{it}]=\tilde{\pi}_i$ $\text{var}[Y_{it}]=\tilde{\pi}_i(1-\tilde{\pi}_i)$ $\text{corr}[Y_{it},Y_{i,t+\ell}]=\rho_\ell$ $=\begin{cases}\rho & \text{for } \ell=1\\ 0 & \text{otherwise,}\end{cases}$
EQC	$Y_{i0}\sim bin(\tilde{\pi}_i)$ $Pr[Y_{it}=1\|y_{i0}]=\tilde{\pi}_i+\rho(y_{i0}-\tilde{\pi}_i), t=1,\ldots$	$E[Y_{it}]=\tilde{\pi}_i$ $\text{var}[Y_{it}]=\tilde{\pi}_i(1-\tilde{\pi}_i)$ $\text{corr}[Y_{it},Y_{i,t+\ell}]=\rho_\ell$ $=\rho^2$

For convenience, we summarize the means, variances, and correlations for all three binary stationary correlation models, as in Table 7.2.

Range For Correlation Index Parameter ρ

The conditional probabilities for AR(1) (7.35) and EQC (7.57) models are similar, therefore by similar calculations as for the AR(1) model, one may show that ρ in the EQC model (7.57) satisfies the same range restriction

$$\max_i\left[-\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}, -\frac{1-\tilde{\pi}_i}{\tilde{\pi}_i}\right]\leq\rho\leq 1, \tag{7.62}$$

as in (7.38).

7.3.4 *Complexity in Likelihood Inferences Under Stationary Binary Correlation Models*

The marginal likelihood estimation by (7.6) is done by ignoring the correlations. If the longitudinal correlation model such as binary AR(1) (7.35), MA(1) (7.45), or EQC (7.57) is known, one may then attempt to obtain the likelihood estimates of β and ρ, by maximizing the likelihood function

$$L(\beta,\rho)=\Pi_{i=1}^{K}[f(y_{i1})\Pi_{t=2}^{T}f(y_{it}|y_{i,t-1})], \tag{7.63}$$

where

$$f(y_{i1}) = \tilde{\pi}_i^{y_{i1}}[1-\tilde{\pi}_i]^{1-y_{i1}}$$

is the binary density with

$$\tilde{\pi}_i = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)},$$

and the conditional density has the form

$$f(y_{it}|y_{i,t-1}) = [\lambda_{it}^*(\beta,\rho|y_{i,t-1})]^{y_{it}}[1-\lambda_{it}^*(\beta,\rho|y_{i,t-1})]^{1-y_{i1}} \quad (7.64)$$

where the conditional probability $\lambda_{it}^*(\beta,\rho|y_{i,t-1}) = P[y_{it}=1|y_{i,t-1}]$ has the forms (7.35), (7.45), and (7.57) under the AR(1), MA(1), and EQC models, respectively. Note that because d_{it} and $d_{i,t-1}$ in (7.45) follow an unknown distribution, one cannot compute the likelihood function (7.64) and hence cannot obtain likelihood estimates for the parameters involved. Further note that even if the distributions of d_{it} and $d_{i,t-1}$ were known such as normal, the integrations would be complicated. A similar integration problem arises in computing the likelihood function under the EQC model (7.57). One of the other major problems with the likelihood approach for the longitudinal binary data analysis is that the model itself for the data may not be known.

Unlike the likelihood approach, the following GQL approach does not need an assumption for any specific model. All that is needed is to assume that the repeated stationary binary data follow any of the three (7.35), (7.45), or (7.57) or similar autocorrelation models. In the nonstationary case one, however, is required to identify the model. Nevertheless, the GQL approach will be much easier than the likelihood approach.

7.3.5 GQL Estimation Approach

Note that the correlation structures for the AR(1), MA(1), and EQC models are given by (7.37), (7.51), and (7.61), respectively. Further note that all three correlation structures may be represented by an autocorrelation matrix of the form:

$$C_i^*(\rho) = (c_{i,(t-\ell)t}^*(\rho)) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix}, \quad (7.65)$$

which is same as (6.25), written for three stationary correlation models, for Poisson longitudinal data.

Let

$$\pi_i = \tilde{\pi}_i 1_T$$

be the mean vector of $y_i = [y_{i1}, \ldots, y_{it}, \ldots, y_{iT}]'$ where 1_T is the T-dimensional unit vector. Because the correlation matrix $C_i^*(\rho)$ in (7.65) represents the correlations of all three models, namely AR(1), MA(1), and EQC, and because this matrix is the same as the correlation matrix in (6.25) under the Poisson model, one may follow the Poisson case and write the GQL estimating equation and the asymptotic variance of the estimates as follows.

GQL Estimating Equation for β

$$\sum_{i=1}^{K} \frac{\partial \pi_i'}{\partial \beta} \Sigma_i^{*-1}(\rho)(y_i - \pi_i) = \sum_{i=1}^{K} X_i' A_i \Sigma_i^{*-1}(\rho)(y_i - \pi_i) = 0, \tag{7.66}$$

where $\Sigma_i^*(\rho) = A_i^{1/2} C_i^*(\rho) A_i^{1/2}$. Note that irrespective of the model, the lag ℓ ($\ell = 1, \ldots, T-1$,) correlation involved in $C_i^*(\rho)$ may be estimated by

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \tilde{y}_{it} \tilde{y}_{i,t+\ell} / K(T-\ell)}{\sum_{i=1}^{K} \sum_{t=1}^{T} \tilde{y}_{it}^2 / KT}, \tag{7.67}$$

where $\tilde{y}_{it}$ is the standardized residual, defined as $\tilde{y}_{it} = (y_{it} - \pi_{it})/\{\sigma_{itt}\}^{1/2}$. Here, unlike in the Poisson case (see Section 6.4.2), $\sigma_{itt} = \tilde{\pi}_i[1 - \tilde{\pi}_i]$ for all $t = 1, \ldots, T$, and A_i in (7.66), has the formula:

$$A_i = \text{diag}[\sigma_{i11}, \ldots, \sigma_{itt}, \ldots, \sigma_{iTT}) = \tilde{\pi}_i[1 - \tilde{\pi}_i] I_T. \tag{7.68}$$

The $p \times T$ covariate matrix X_i' is the same as in the Poisson case and is given by

$$X_i' = [\tilde{x}_i, \ldots, \tilde{x}_i] \text{ with } \tilde{x}_i = (x_{i1}, \ldots, x_{ip})'.$$

'Working' Independence Assumption Based GQL Estimation

By using $C_i^*(\rho) = I_T$ in (7.65), a 'working' independence assumption based GQL estimate may be obtained by solving

$$\sum_{i=1}^{K} X_i' A_i \Sigma_i^{*-1}(\rho)(y_i - \pi_i)|_{\rho=0} = \sum_{i=1}^{K} X_i'(y_i - \pi_i) = 0, \tag{7.69}$$

which is also referred to as the 'working' independence assumption based GEE estimate [Liang and Zeger (1986)].

7.3.5.1 Efficiency of the Independence Assumption Based Estimation

Let $\hat{\beta}_{GQL}$ and $\hat{\beta}(I)$, be the solutions of (7.66) and (7.69), respectively. The asymptotic variances of these estimators have the same forms as in the Poisson case, and they are given by (6.29) and (6.31), respectively. There are, however, two significant differences in the formulas between the count and binary cases.

Remark 1. First, the A_i matrix has the formula given by (7.68); that is, $A_i = \tilde{\pi}_i(1-\tilde{\pi}_i)I_T$ in the present binary case, whereas in the Poisson case $A_i = \tilde{\mu}_i I_T$.

Remark 2. Second, the corresponding elements of the $C_i^*(\rho)$ matrix under the Poisson MA(1) and EQC structures are different from those of the binary structures. For example, the Poisson EQC structure (6.20) produces correlations ρ (6.22), whereas the binary EQC structure (7.57) yields the correlations as ρ^2 (7.61). These differences have to be taken into account while computing the asymptotic variances of both GQL and 'working' independence based GQL estimators.

Note that to compute the relative efficiencies in the binary case, one also has to pay attention to the ranges for the correlation index parameter ρ, which are different from the Poisson case. In the present binary case, the ranges for ρ, for example, under the stationary AR(1) and EQC models are given by (7.38) and (7.62), respectively. For the purpose, one needs to compute the value of $\tilde{\pi}_i$ for all $i = 1,\ldots,K$. For example, for $K = 100$, and $\beta_1 = \beta_2 = 1.0$, and the covariates

$$x_{it1} = \tilde{x}_{i1} = 1.0, \ \text{for all } i = 1,\ldots,K, \ \text{and } \ t = 1,\ldots,T,$$

and

$$x_{it2} = \tilde{x}_{i2} = \begin{cases} -1 & \text{for } t = 1,\ldots,T;\ i = 1,\ldots,K/4 \\ 0 & \text{for } t = 1,\ldots,T;\ i = (K/4)+1,\ldots,3K/4 \\ 1 & \text{for } t = 1,\ldots,T;\ i = (3K/4)+1,\ldots,K, \end{cases}$$

we compute the values of $\tilde{\pi}_i$ by

$$\tilde{\pi}_i = \frac{\exp(\tilde{x}_{i1}\beta_2 + \tilde{x}_{i2}\beta_2)}{1+\exp(\tilde{x}_{i1}\beta_2 + \tilde{x}_{i2}\beta_2)},$$

and they are found to be

$$\tilde{\pi}_i = \begin{cases} 0.50 & \text{for } i = 1,\ldots,K/4 \\ 0.73 & \text{for } i = (K/4)+1,\ldots,3K/4 \\ 0.88 & \text{for } i = (3K/4)+1,\ldots,K. \end{cases}$$

These values of probabilities for all $K = 100$ individuals produce

$$\max_i\left[-\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}, -\frac{1-\tilde{\pi}_i}{\tilde{\pi}_i}\right] = -0.14,$$

yielding the valid range for ρ as $-0.14 < \rho < 1.0$ under the AR(1) and EQC models with the aforementioned covariates.

Now for selected values of ρ within the range $-0.14 < \rho < 1.0$ for AR(1) and EQC models, for example, and following the above Remarks 1 and 2, we compute the efficiencies of the $\hat{\beta}_I$ to $\hat{\beta}_{GQL}$ by using the formulas (6.29) and (6.31) in (6.32) in Section 6.4. These efficiencies of the independence based estimator with regard to binary AR(1) and EQC models are shown in Table 7.3, along with the efficiencies of certain GEE estimators. It is clear that $\hat{\beta}_I$ always has larger or equal asymptotic variances as compared to the GQL estimator $\hat{\beta}_{GQL}$, showing that the GQL estimator is always more efficient than the 'working' independence based estimator, irrespective of the discrete nature of the data whether they are binary or count. Note that when these results of Tables 7.2 and 7.3 are compared to those of Table 6.2 for the Poisson case, the efficiency of $\hat{\beta}_I$ remains the same for the Poisson or binary AR(1)-type data, whereas the efficiency of $\hat{\beta}_I$ gets worse under the binary EQC model as compared to the Poisson EQC model.

7.3.6 GEE Approach and Its Limitations for Binary Data

For the estimation of the regression effects by using the GEE approach [Liang and Zeger (1986)] for the Poisson data, we refer to Section 6.4.3. To be specific, the regression effect β is estimated by solving the GEE given in (6.33) and such GEE based estimator $\hat{\beta}_G$ has the asymptotic covariance matrix given by (6.34). Note that these formulas for the count data may still be used for the binary data case, except that μ_i vector is replaced by $\pi_i = 1'_T \tilde{\pi}_i$ and the A_i matrix now has the form

$$A_i = \text{diag}[\sigma_{i11}, \ldots, \sigma_{itt}, \ldots, \sigma_{iTT}) = \tilde{\pi}_i[1 - \tilde{\pi}_i]I_T$$

as in (7.68). As far as the working correlation matrices are concerned, the formulas for $R(\alpha)$ and $R(\alpha_0)$ remain the same as in the Poisson data case. Here α is the so-called 'working' correlation parameter and its moment estimate, say $\hat{\alpha}$, converges in limit to α_0. The values of α_0 corresponding to the true correlation index parameter are calculated in a similar fashion as in Section 6.4.3 [see also Sutradhar and Das (1999)].

The purpose of this section is to examine whether $\hat{\beta}_G$ can always be more efficient than the 'working' independence based estimator $\hat{\beta}_I$ under the stationary correlation models for the binary data. Recall that following Sutradhar and Das (1999), it was demonstrated in Section 6.4.3 for the stationary Poisson data that $\hat{\beta}_G$, in fact, can be less efficient than the simpler $\hat{\beta}_I$. Thus the GEE based estimation cannot be trusted. Note that Sutradhar and Das (1999) have examined this efficiency issue for a nonstationary binary dataset following stationary binary correlations, which was earlier used by Liang and Zeger (1986). Here we consider stationary binary data with means, variances, and correlations, free from time-dependent covariates. For simplicity, we now consider two models, namely AR(1) (7.35) and EQC (7.57) with true correlation index parameter $\rho = 0.3,\ 0.7$ which are within the valid range for the covariates considered in Section 7.3.5.1. The efficiencies of GEE estimators are computed by using (6.35) under the assumption that true data follow AR(1) model (7.35) whereas one uses the EQC model (7.57) as the 'working' correlation model, and vice versa. The results are given in Table 7.3. As mentioned in Section 7.3.5.1, the same table also contains the efficiencies of the independence based estimator $\hat{\beta}_I$ as compared to the GQL (true model based) estimator $\hat{\beta}_{GQL}$.

Table 7.3 Percentage relative efficiency of $\hat{\beta}_I$ and $\hat{\beta}_G$ to the GQL estimator $\hat{\beta}_{GQL}$ with true correlation matrix $C(\rho)$ for AR(1) and EQC (exchangeable) structure, for $\pi_{it} = \tilde{\pi}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)/[1+\exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)]$ with $\beta_1 = \beta_2 = 1$.

	Working/True Correlation Structures											
	AR(1)/EQC						EQC/AR(1)					
T	ρ	$\hat{\beta}_{1I}$	$\hat{\beta}_{2I}$	α_0	$\hat{\beta}_{1G}$	$\hat{\beta}_{2G}$	ρ	$\hat{\beta}_{1I}$	$\hat{\beta}_{2I}$	α_0	$\hat{\beta}_{1G}$	$\hat{\beta}_{2G}$
5	0.3	100	100	0.49	93	93	0.3	98	98	0.15	98	98
	0.7	100	100	0.83	90	90	0.7	95	95	0.52	95	95
10	0.3	100	100	0.65	87	87	0.3	99	99	0.08	99	99
	0.7	100	100	0.90	88	87	0.7	93	93	0.35	93	93
15	0.3	100	100	0.74	83	83	0.3	99	99	0.06	99	99
	0.7	100	100	0.93	85	85	0.7	93	93	0.26	93	93

The results from Table 7.3 show that the independence assumption based estimator $\hat{\beta}_I$ is equally efficient to the GQL estimator $\hat{\beta}_{GQL}$ when the true correlation structure is EQC. This $\hat{\beta}_I$ is less efficient than $\hat{\beta}_{GQL}$ when binary data follow the AR(1) correlation structure. Thus, $\hat{\beta}_{GQL}$ is always the same or more efficient than the independence based estimator. These results are exactly the same as in the Poisson case (see Table 6.2).

Next, when the GEE estimator $\hat{\beta}_G$ is compared to the independence based estimator $\hat{\beta}_I$, the former appears to be less efficient when binary data follow the EQC model but estimation is done based on the AR(1) model. When the data follow the AR(1) model but estimation is done based on the EQC model, the independence based estimator $\hat{\beta}_I$ appears to be equally efficient to the GEE based estimator $\hat{\beta}_G$. Thus, this comparison along with the comparison made in Section 6.4.3 for the stationary Poisson longitudinal data clearly demonstrates that the so-called GEE approach cannot be trusted as it may fail to produce more efficient estimates than the independence assumption based estimation approach.

7.4 Inferences in Nonstationary Correlation Models for Repeated Binary Data

In this section, following Sutradhar (2010), we provide a generalization to the three stationary binary correlation models discussed in Section 7.3. In the nonstationary case, it is not only that the marginal means and variances are nonstationary, i.e., they are functions of the time-dependent covariates, the correlations also are nonstationary as they become the functions of time dependent covariates. Note that it is reasonable to expect that ignoring the nonstationary correlations may have adverse effects on the estimation of the correlation index parameter that one uses to define the stationary correlation structure. This in turn may cause efficiency loss for the estimates of the regression effects. As a remedy, it then becomes an issue to identify the nonstationary correlation structure within the autocorrelation class of models. As shown in Section 7.4.5, the identification is done by computing the lag correlations $\hat{\rho}_\ell$ $(\ell = 1, \ldots, T-1)$ for $\beta = 0$, and matching their pattern with that of the expected values under a given nonstationary model. Remark that for simplicity, one may identify the nonstationary correlation structure by comparing the pattern of $\hat{\rho}_\ell$ $(\ell = 1, \ldots, T-1)$ with that of stationary correlation models. Once the identification is done, one goes back to the identified model and does the GQL estimation for the regression effects to obtain consistent and efficient estimates. This GQL estimation, similar to the Poisson case (Section 6.5.2), is given briefly in Section 7.4.4.

We now provide three binary nonstationary correlation models, namely AR(1), MA(1), and EQC, as a generalization of the stationary binary AR(1), MA(1), and EQC models, given in Sections 7.3.1, 7.3.2, and 7.3.3, respectively. Note that all three correlation models given below produce the same time-dependent marginal means and variances, their correlation structures being different and they are functions of time-dependent covariates.

7.4.1 Nonstationary AR(1) Correlation Model

Suppose that the repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are generated following the probability relationship

$$\begin{aligned} Pr[Y_{i1} = 1] &= \pi_{i1} \\ Pr[Y_{it} = 1 | y_{i,t-1}] &= \pi_{it} + \rho(y_{i,t-1} - \pi_{i,t-1}), \text{ for } t = 2, \ldots, T, \end{aligned} \tag{7.70}$$

where $\pi_{it} = \exp(x_{it}'\beta)/[1+\exp(x_{it}'\beta)]$, for all $t = 1, \ldots, T$. This model clearly yields the means and the variances as

$$\begin{aligned} E(Y_{it}) &= \pi_{it} \\ \text{var}(Y_{it}) &= \sigma_{itt} = a_{itt} = \pi_{it}(1-\pi_{it}), \end{aligned} \tag{7.71}$$

for $t = 1, \ldots, T$.

Next, for $u < t$, by using the model relationship (7.70), one may compute the covariance between y_{iu} and y_{it} as

$$\begin{aligned}
\mathrm{cov}[Y_{iu}, Y_{it}] &= E[Y_{iu}Y_{it}] - E[Y_{iu}]E[Y_{it}] \\
&= E_{y_{iu}} Y_{iu} E_{y_{i,t-(t-u-1)}}[\ldots[E_{y_{i,t-2}}[E_{y_{i,t-1}}[Y_{it}|y_{i,t-1}]|y_{i,t-2}]\ldots]|y_{i,t-(t-u-1)}] \\
&\quad -\pi_{iu}\pi_{it} \\
&= E_{y_{iu}} Y_{iu}[\pi_{it} + \sum_{j=1}^{t-u-1} \rho^j \pi_{i,t-j} + \rho^{t-u}(Y_{iu} - \pi_{iu}) - \sum_{j=1}^{t-u-1} \rho^j \pi_{i,t-j}] - \pi_{iu}\pi_{it} \\
&= \pi_{iu}\pi_{it} + \rho^{t-u} E_{y_{iu}}[Y_{iu}(Y_{iu} - \pi_{iu})] - \pi_{iu}\pi_{it} \\
&= \rho^{t-u}\pi_{iu}[1 - \pi_{iu}] \\
&= \rho^{t-u}\sigma_{iuu}. \qquad (7.72)
\end{aligned}$$

Consequently, for all $u, t = 1, \ldots, T$, the nonstationary correlation matrix is given by

$$c_{i,u,t}^{(ns)}(x_{iu}, x_{it}, \rho) = \mathrm{corr}(Y_{iu}, Y_{it}) = \begin{cases} \rho^{t-u}\left[\frac{\sigma_{iuu}}{\sigma_{itt}}\right]^{1/2}, & \text{for } u < t \\ \rho^{u-t}\left[\frac{\sigma_{itt}}{\sigma_{iuu}}\right]^{1/2}, & \text{for } u > t. \end{cases} \qquad (7.73)$$

Note that in (7.73), the correlations are nonstationary. This is because σ_{itt}, for example, depends on x_{it}. Further note that the ρ parameter in (7.73) must satisfy the range restriction

$$\max\left[-\frac{\pi_{it}}{1 - \pi_{i,t-1}}, -\frac{1 - \pi_{it}}{\pi_{i,t-1}}\right] \le \rho \le \min\left[\frac{1 - \pi_{it}}{1 - \pi_{i,t-1}}, \frac{\pi_{it}}{\pi_{i,t-1}}\right]. \qquad (7.74)$$

Remarks on the Stationary Correlation Structure: Under the stationary case, $\sigma_{itt} = \tilde{\pi}_i(1 - \tilde{\pi}_i)$ for all $t = 1, \ldots, T$, therefore the nonstationary correlation structure in (7.73) reduces to the same form $c_{i,ut}^*(\rho) = \rho^{|t-u|}$ as in (7.37), which is also the AR(1) stationary correlation structure for the longitudinal count data [see (6.17)]. Thus, even though the nonstationary AR(1) correlation structures for the count (6.46) and binary (7.73) data are different, their stationary correlation structures are, however, the same. Thus, in any inferences for the regression effects under the stationary case, one can use the unique autocorrelation structure (7.65) irrespective of the situations whether longitudinal data are count or binary, also irrespective of the processes whether AR(1), MA(1), or EQC, but in the nonstationary case it is not only that the correlation structures are different for binary and count data, they are also different under different processes.

7.4.2 Nonstationary MA(1) Correlation Model

Suppose that the repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are generated following the probability relationship

$$\begin{aligned} Pr[Y_{i1} = 1] &= \pi_{i1} \\ Pr[Y_{it} = 1 | d_{it}, d_{i,t-1}] &= d_{it} + \rho d_{i,t-1}, \text{ for } t = 2, \ldots, T, \end{aligned} \tag{7.75}$$

where the d_{it}s are independently distributed with mean ξ^*_{it} and variance η_{it}; that is,

$$d_{it} \overset{id}{\sim} \left[\xi^*_{it} = \sum_{j=0}^{t-1} (-\rho)^j \pi_{i,t-j},\ \eta_{it} = [\frac{\sum_{j=0}^{t-1}(-\rho)^j \pi_{i,t-j}}{\sum_{j=0}^{t-1}(-\rho)^j}][1 - \frac{\sum_{j=0}^{t-1}(-\rho)^j \pi_{i,t-j}}{\sum_{j=0}^{t-1}(-\rho)^j}] \right],$$

for all $t = 1, \ldots, T$. Under this model, one obtains the means given by

$$\begin{aligned} E[Y_{i1}] &= \pi_{i1} \\ E[Y_{it}] &= E_{d_{it}, d_{i,t-1}} E[Y_{it} | d_{it}, d_{i,t-1}] \\ &= E_{d_{it}, d_{i,t-1}} [d_{it} + \rho d_{i,t-1}] \\ &= \left[\sum_{j=0}^{t-1} (-\rho)^j \pi_{i,t-j} \right] + \rho \left[\sum_{j=0}^{t-2} (-\rho)^j \pi_{i,t-1-j} \right] \\ &= \pi_{it}, \text{ for } t = 2, \ldots, T, \end{aligned} \tag{7.76}$$

and the variances, by similar calculations as in (7.48), given by

$$\begin{aligned} \text{var}[Y_{i1}] &= \pi_{i1}[1 - \pi_{i1}] \\ \text{var}[Y_{it}] &= E_{d_{it}, d_{i,t-1}} \text{var}[Y_{it} | d_{it}, d_{i,t-1}] + \text{var}_{d_{it}, d_{i,t-1}} E[Y_{it} | d_{it}, d_{i,t-1}] \\ &= \pi_{it} - \pi_{it}^2 = \pi_{it}[1 - \pi_{it}], \end{aligned} \tag{7.77}$$

respectively.

Next, for $u < t$, by using the model relationship (7.75), and by similar calculations as in (7.49), one obtains

$$\begin{aligned} \text{cov}(Y_{iu}, Y_{it}) &= \text{cov}_{d_{iu}, d_{i,u-1}, d_{it}, d_{i,t-1}} [d_{iu} + \rho d_{i,u-1}, d_{it} + \rho d_{i,t-1}] \\ &= \rho \left[[\frac{\sum_{j=0}^{u-1}(-\rho)^j \pi_{i,u-j}}{\sum_{j=0}^{u-1}(-\rho)^j}][1 - \frac{\sum_{j=0}^{u-1}(-\rho)^j \pi_{i,u-j}}{\sum_{j=0}^{u-1}(-\rho)^j}] \right], \end{aligned} \tag{7.78}$$

yielding the nonstationary lag $|t - u|$ correlations as

$$c^{(ns)}_{i,u,t}(x_{iu}, x_{it}, \rho) = corr(Y_{iu}, Y_{it}) = \begin{cases} \frac{\rho\left[[\frac{\sum_{j=0}^{u-1}(-\rho)^j \pi_{i,u-j}}{\sum_{j=0}^{u-1}(-\rho)^j}][1-\frac{\sum_{j=0}^{u-1}(-\rho)^j \pi_{i,u-j}}{\sum_{j=0}^{u-1}(-\rho)^j}]\right]}{\left[\sqrt{\pi_{it}(1-\pi_{it})}\sqrt{\pi_{iu}(1-\pi_{iu})}\right]} & \text{for } t-u=1 \\ 0 & \text{for } (t-u)>1 \end{cases} \tag{7.79}$$

Remarks on Stationary Correlation Structure

For the stationary correlated data, the nonstationary correlation structure in (7.79) reduces to the correlation structure in (7.51), which has the form

$$c^*_{i,ut}(\rho) = \text{corr}(Y_{iu}, Y_{it}) = \begin{cases} \rho & \text{for } |u-t| = 1 \\ 0 & \text{otherwise,} \end{cases}$$

Note that even though this stationary correlation structure is the same as for the Poisson case (6.19), the lag 1 correlation formula for the Poisson case is $c^*_{i,u,u+1} = \rho/1+\rho$ which is different from the present binary case. This is expected as the correlation index parameter ρ for the count and binary cases has a different interpretation. Similar things happen also under the EQC structure for count and binary data.

7.4.3 Nonstationary EQC Model

Suppose that y_{i0} is an unobservable initial binary response with its mean the same as that of y_{i1}. Also, all repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are generated following the probability relationship

$$Pr[Y_{it} = 1|y_{i0}] = \pi_{it} + \rho(y_{i0} - \pi_{i1}), \;\; \text{for } t = 1, \ldots, T. \tag{7.80}$$

It then follows that

$$\begin{aligned} E[Y_{it}] &= E_{y_{i0}} E[Y_{it}|y_{i0}] = E_{y_{i0}}[Pr(Y_{it} = 1|y_{i0})] \\ &= E_{y_{i0}}[\pi_{it} + \rho(y_{i0} - \pi_{i1})] = \pi_{it}, \end{aligned} \tag{7.81}$$

and

$$\begin{aligned} \text{var}[Y_{it}] &= E_{y_{i0}}[\text{var}\{Y_{it}|y_{i0}\}] + \text{var}_{y_{i0}}[E\{Y_{it}|y_{i0}\}] \\ &= E_{y_{i0}}[\{\pi_{it} + \rho(y_{i0} - \pi_{i1})\}(1 - \{\pi_{it} + \rho(y_{i0} - \pi_{i1})\})] \\ &\quad + \text{var}_{y_{i0}}[\pi_{it} + \rho(y_{i0} - \pi_{i1})] \\ &= E_{y_{i0}}\{\pi_{it} + \rho(y_{i0} - \pi_{i1})\} - [E_{y_{i0}}\{\pi_{it} + \rho(y_{i0} - \pi_{i1})\}]^2 \\ &= \pi_{it}(1 - \pi_{it}). \end{aligned} \tag{7.82}$$

These means (7.81) and the variances (7.82) under the EQC model are same as that for the AR(1) binary process given in (7.71), and also for the MA(1) binary process given in (7.76) – (7.77).

For $u \neq t$, by using the model relationship (7.80), one obtains

$$\begin{aligned}
\text{cov}(Y_{iu}, Y_{it}) &= E_{y_{i0}} \text{cov}[(Y_{iu}, Y_{it})|y_{i0}] \\
&\quad + \text{cov}_{y_{i0}} [(\pi_{iu} + \rho(y_{i0} - \pi_{i1})), (\pi_{it} + \rho(y_{i0} - \pi_{i1}))] \\
&= \text{cov}_{y_{i0}} [(\rho(y_{i0} - \pi_{i1})), (\rho(y_{i0} - \pi_{i1}))] \\
&= \left[\rho^2 \pi_{i0}(1 - \pi_{i0})\right] \\
&\equiv \left[\rho^2 \pi_{i1}(1 - \pi_{i1})\right],
\end{aligned} \tag{7.83}$$

yielding the correlations as

$$c_{i,ut}^{(ns)}(x_{iu}, x_{it}, \rho) = \text{corr}(Y_{iu}, Y_{it}) = \frac{\left[\rho^2 \pi_{i1}(1 - \pi_{i1})\right]}{\left[\sqrt{\pi_{it}(1 - \pi_{it})}\sqrt{\pi_{iu}(1 - \pi_{iu})}\right]}. \tag{7.84}$$

Remarks on Stationary Correlation Structure: Note that for the stationary case, the correlations in (7.84) reduce to the form

$$c_{i,ut}^{*}(\rho) = \rho^2,$$

which is similar to (7.61). Here for the stationary EQC binary data, ρ^2 represents the constant correlation, whereas ρ is the constant correlation parameter under the stationary EQC model for count data. But this difference in parameter selection does not cause any problems in inferences for the regression parameters. This is because the constant stationary correlation matrix form (7.65) is used any way in the GQL estimating equation (7.66), for all stationary cases.

For convenience, we summarize the nonstationary binary AR(1), MA(1), and EQC models along with their correlation structures in Table 7.4.

7.4.4 Nonstationary Correlations Based GQL Estimation

GQL Estimating Equation for β: Similar to the GQL estimation (6.56) under nonstationary Poisson models, we now solve the GQL estimating equation for β given by

$$\sum_{i=1}^{K} \frac{\partial \pi_i'}{\partial \beta} \Sigma_i^{(ns)^{-1}}(\hat{\rho})(y_i - \pi_i) = 0, \tag{7.85}$$

where

$$\pi_i = (\pi_{i1}, \cdots, \pi_{it}, \cdots, \pi_{iT})' \text{ and } \Sigma_i^{(ns)}(\hat{\rho}) = A_i^{1/2} C_i^{(ns)}(x_i, \hat{\rho}) A_i^{1/2}$$

Table 7.4 A class of nonstationary correlation models for longitudinal binary data and basic properties.

Model	Dynamic Relationship	Mean, Variance, & Correlations
AR(1)	$Pr[Y_{i1}=1]=\pi_{i1}$ $Pr[Y_{it}=1 \mid y_{i,t-1}]=\pi_{it}+\rho(y_{i,t-1}-\pi_{i,t-1}),\ t=2,\ldots,T$	$E[Y_{it}]=\pi_{it}$ $\text{var}[Y_{it}]=\pi_{it}(1-\pi_{it})$ $\text{corr}[Y_{iu},Y_{it}]=\rho^{(ns)}_{\vert u-t\vert}$ $=\rho^{\vert t-u\vert}\left[\frac{\sigma_{iuu}}{\sigma_{itt}}\right]^{\frac{1}{2}}$
MA(1)	$Pr[Y_{i1}=1]=\pi_{i1}$ $Pr[Y_{it}=1 \mid d_{it},d_{i,t-1}]=d_{it}+\rho d_{i,t-1},t=2,\ldots,T$ $d_{it}\overset{id}{\sim}\left[\sum_{j=0}^{t-1}(-\rho)^j\pi_{i,t-j},\ [\frac{\sum_{j=0}^{t-1}(-\rho)^j\pi_{i,t-j}}{\sum_{j=0}^{t-1}(-\rho)^j}][1-\frac{\sum_{j=0}^{t-1}(-\rho)^j\pi_{i,t-j}}{\sum_{j=0}^{t-1}(-\rho)^j}]\right]$ $\text{corr}[Y_{iu},Y_{it}]=\rho^{(ns)}_{\vert u-t\vert}$ $=\begin{cases}\frac{\rho\left[[\frac{\sum_{j=0}^{u-1}(-\rho)^j\pi_{i,u-j}}{\sum_{j=0}^{u-1}(-\rho)^j}][1-\frac{\sum_{j=0}^{u-1}(-\rho)^j\pi_{i,u-j}}{\sum_{j=0}^{u-1}(-\rho)^j}]\right]}{\left[\sqrt{\pi_{it}(1-\pi_{it})}\sqrt{\pi_{iu}(1-\pi_{iu})}\right]} & \text{for } \vert u-t\vert=1\\ 0 & \text{otherwise,}\end{cases}$	$E[Y_{it}]=\pi_{it}$ $\text{var}[Y_{it}]=\pi_{it}(1-\pi_{it})$
EQC	$Pr[Y_{it}=1 \mid y_{i0}]=\pi_{it}+\rho(y_{i0}-\pi_{i1}),t=1,\ldots,T$ $Y_{i0}\sim bin(\pi_{i1})$	$E[Y_{it}]=\pi_{it}$ $\text{var}[Y_{it}]=\pi_{it}(1-\pi_{it})$ $\text{corr}[Y_{iu},Y_{it}]=\rho^{(ns)}_{\vert u-t\vert}$ $=\frac{\rho^2\pi_{i1}(1-\pi_{i1})}{\sqrt{\pi_{iu}(1-\pi_{iu})\pi_{it}(1-\pi_{it})}}$

are the mean vector and true covariance matrix of $y_i=(y_{i1},\cdots,y_{it},\cdots,y_{iT})'$. Note that

$$E[Y_{it}]=\pi_{it}=\frac{\exp(x_{it}'\beta)}{1+\exp(x_{it}'\beta)}$$

and

$$A_i=\text{diag}[\sigma_{i11},\ldots,\sigma_{itt},\ldots,\sigma_{iTT}],$$

with $\sigma_{itt}=\pi_{it}(1-\pi_{it})$, remain the same under all three nonstationary binary AR(1) $(7.70-7.71)$, MA(1) $(7.75-7.77)$, and EQC $(7.80-7.82)$ models, but their correlation structures, that is, $C_i^{(ns)}(x_i,\rho)$, given in (7.73) under the AR(1) model, (7.79) under the MA(1) model, and in (7.84) under the EQC model, are different from each other. For this reason, it is necessary to identify the correlation structure (see the next section). Once the correlation structure is identified, by using $\partial\pi_i'/\partial\beta=X_i'A_i$ in (7.85), one may solve the GQL estimating equation

$$\sum_{i=1}^{K}X_i'A_i\Sigma_i^{(ns)^{-1}}(\hat{\rho})(y_i-\pi_i)=0, \tag{7.86}$$

for β. Similar to the Poisson case, let $\hat{\beta}_{GQL}$ denote the solution of (7.86) after using $\hat{\rho}$ computed under the selected model. Under mild regularity conditions one may then show that $\hat{\beta}_{GQL}$ has the asymptotic (as $K\to\infty$) normal distribution given by

$$K^{\frac{1}{2}}(\hat{\beta}_{GQL} - \beta) \sim N\left(0,\, K\left[\sum_{i=1}^{K} X_i' A_i \Sigma_i^{(ns)^{-1}} A_i X_i\right]^{-1}\right). \qquad (7.87)$$

We now show how to compute $\hat{\rho}$ under all three binary models.

7.4.4.1 Estimation of ρ Parameter Under Binary AR(1) Model

Moment Equation for ρ: Recall that the correlations under the nonstationary binary AR(1) model (7.70) are given in (7.73). Now to estimate the correlation index parameter ρ in this correlation structure (7.73), one may use the moment estimate of ρ given by

$$\hat{\rho} = \frac{\sum_{i=1}^{K}\sum_{t=2}^{T} \tilde{y}_{it}\tilde{y}_{i,t-1}}{\sum_{i=1}^{K}\sum_{t=1}^{T} \tilde{y}_{it}^2} \frac{KT}{\sum_{i=1}^{K}\sum_{t=2}^{T}[\sigma_{i,t-1,t-1}/\sigma_{itt}]^{\frac{1}{2}}}, \qquad (7.88)$$

where $\tilde{y}_{it} = [y_{it} - \pi_{it}]/\sqrt{\sigma_{itt}}$, with $\sigma_{itt} = \pi_{it}(1-\pi_{it})$. Note that the formula for ρ given by (7.88) was obtained by equating the lag 1 sample autocorrelation with its population counterpart given by (7.73). Furthermore, $\hat{\rho}$ computed by (7.88) must satisfy the range restriction given in (7.74). This implies that if the value of $\hat{\rho}$ computed by (7.88) falls beyond the range shown in (7.74), we use the upper limit of ρ given in (7.74) as the estimate of ρ.

7.4.4.2 Estimation of ρ Parameter Under Binary MA(1) Correlation Model

Similar to the Poisson MA(1) model, the formula for lag 1 correlations given by (7.79) under the nonstationary MA(1) model (7.75) involves a complicated summation. Thus, it is convenient to solve the moment equation for ρ by using the Newton−Raphson iterative technique. To be specific, by writing the moment equation as

$$g(\rho) = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-1} \tilde{y}_{it}\tilde{y}_{i,t+1}/K(T-1)}{\sum_{i=1}^{K}\sum_{t=1}^{T} \tilde{y}_{it}^2/KT} - \frac{\rho}{T-1}\sum_{u=1}^{T-1}\left[\frac{[\frac{\sum_{j=0}^{u-1}(-\rho)^j \pi_{i,u-j}}{\sum_{j=0}^{u-1}(-\rho)^j}][1 - \frac{\sum_{j=0}^{u-1}(-\rho)^j \pi_{i,u-j}}{\sum_{j=0}^{u-1}(-\rho)^j}]}{\left[\sqrt{\pi_{i,u+1}(1-\pi_{i,u+1})}\sqrt{\pi_{iu}(1-\pi_{iu})}\right]}\right] = 0, \qquad (7.89)$$

we solve for ρ iteratively by using the Newton−Raphson iterative formula

$$\hat{\rho}(r+1) = \hat{\rho}(r) - \left[\{\frac{\partial g(\rho)}{\partial \rho}\}^{-1} g(\rho)\right]_{(r)},$$

where $[\cdot]_{(r)}$ denotes that the expression within brackets is evaluated at $\rho = \hat{\rho}(r)$, the rth iterative value of ρ. Note that $\hat{\rho}$ must satisfy the appropriate range restrictions, which are, however, complicated to derive under the MA(1) model. One of the advantages of using the moment method for the estimation of ρ is that the estimates usually satisfy the underlying restrictions irrespective of the formulas for the ranges.

7.4.4.3 Estimation of ρ Parameter Under Exchangeable (EQC) Correlation Model

The moment estimating equation for the ρ parameter for the exchangeable model is quite similar to that for the AR(1) model. The difference between the two equations is that under the AR(1) process we have considered all lag 1 standardized residuals, whereas under the exchangeable model, one is required to use standardized residuals of all possible lags. Thus, following (6.58) for the AR(1) model, we write the moment formula for ρ under the exchangeable model as

$$\hat{\rho}^2 = \frac{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\tilde{y}_{it}\tilde{y}_{i,t+\ell}}{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\tilde{y}_{it}^2}\,\frac{KT}{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\frac{\pi_{i1}(1-\pi_{i1})}{[\pi_{it}(1-\pi_{it})\pi_{i,t+\ell}(1-\pi_{i,t+\ell})]^{\frac{1}{2}}}}, \tag{7.90}$$

where $\tilde{y}_{it} = [y_{it} - \pi_{it}]/\sqrt{\pi_{it}(1-\pi_{it})}$.

7.4.5 Model Selection

As it was argued in Section 6.5.3 for the model selection for Poisson correlated data that the pattern of lag correlations can be exploited to identify the correlation structures in both stationary (if needed) and nonstationary cases. This argument also holds for the longitudinal binary data. Thus, to select a nonstationary binary correlation model among the three possible AR(1), MA(1), and EQC models, we first compute the estimated lag correlations by using

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\tilde{y}_{it}\tilde{y}_{i,t+\ell}/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}\tilde{y}_{it}^2/KT}, \tag{7.91}$$

[see also (6.61)] where $\tilde{y}_{it} = [y_{it} - \pi_{it}]/\sqrt{[\pi_{itt}(1-\pi_{itt})]}$. This can be done by using $\beta = 0$ in the formulas for π_{it}. These values of $\hat{\rho}_\ell$, which are stationary correlation values, may be enough to identify the correlation structure. For finer identification, one computes the approximate expected values; that is, $E[\hat{\rho}_\ell]$ under all three models for all possible trial values of ρ, the correlation index parameter. Next, the closeness of the pattern for $\hat{\rho}_\ell$ with that of $E[\hat{\rho}_\ell]$ under a model determines the selection of the model.

A first-order approximation to the formulas for $E[\hat{\rho}_\ell]$ under the nonstationary binary AR(1), MA(1), or EQC models is given as follows.

$$\text{For } AR(1): E[\hat{\rho}_\ell] = \frac{\rho^\ell}{K(T-\ell)}$$

$$\times \sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \left[\frac{\sigma_{itt}}{\sigma_{i,t+\ell,t+\ell}} \right]^{\frac{1}{2}} \quad \text{for } \ell = 1, \ldots, T-1 \tag{7.92}$$

$$\text{For } MA(1) : E[\hat{\rho}_\ell] = \begin{cases} \frac{\rho}{K(T-1)} \sum_{i=1}^{K} \sum_{t=1}^{T-1} \frac{1}{\sqrt{\sigma_{itt}\sigma_{i,t+1,t+1}}} \\ \times \left[\{\sum_{j=0}^{t-1} (-\rho)^j \pi_{i,t-j} / \sum_{j=0}^{t-1} (-\rho)^j\} \right. \\ \left. \{1 - \sum_{j=0}^{t-1} (-\rho)^j \pi_{i,t-j} / \sum_{j=0}^{t-1} (-\rho)^j\} \right] \\ 0 \text{ for } \ell = 2, \ldots, T-1, \end{cases} \tag{7.93}$$

$$\text{For } EQC : E[\hat{\rho}_\ell] = \frac{\rho^2}{K(T-\ell)} \times \sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \left[\frac{\sigma_{i11}}{\{\sigma_{itt}\sigma_{i,t+\ell,t+\ell}\}^{\frac{1}{2}}} \right] \quad \text{for } \ell = 1, \ldots, T-1, \tag{7.94}$$

with $\sigma_{itt} = \pi_{it}(1-\pi_{it})$, where $\pi_{it} = \exp(x'_{it}\beta)/[1+\exp(x'_{it}\beta)]$ for all $t = 1, \ldots, T$.

7.5 SLID Data Example

7.5.1 Introduction to the SLID Data

The Survey of Labour and Income Dynamics (SLID) is a longitudinal household survey, designed to capture changes in the economic well-being of Canadians over time. Statistics Canada has conducted this survey from 1993 to 1998. In this study, we considered all SLID longitudinal respondents who were either employed or unemployed during 1993 – 1996. Those who were out of the labour force for at least a part of the year were excluded. A binary response variable 'unemployed all year', derived from a variable 'labour force status for the year', assigns value $y_{it} = 1$ to the ith individual who was unemployed for the full year 't'', and $y_{it} = 0$ if the individual was employed for the full year 't'' or a part of year employed and a part unemployed.

Note that the SLID data were earlier analyzed by Sutradhar and Kovacevic (2000) and Sutradhar, Rao, and Pandit (2008), among others. Sutradhar and Kovacevic (2000) have, however, considered an ordinal longitudinal multinomial model for the jobless spell response variable and studied the complete responses from 16,890 individuals collected over a period of two years 1993 and 1994, whereas, Sutradhar, Rao, and Pandit (2008) have analyzed the SLID data for four years from 1993 to 1996 collected from 15,731 individuals, by fitting a nonlinear binary dynamic mixed model. This nonlinear mixed model for binary data is discussed in detail in Chapter 9. Here, it is of interest to illustrate an application of the nonstationary LDCP (linear dynamic conditional probability) models for binary data discussed in Section 7.4. For the purpose, we consider the same dataset used by Sutradhar, Rao, and Pandit (2008). Thus, in our notation, $K = 15,731$ and $T = 4$. The frequency distribution for the 'unemployment' status for these 15,731 individuals over four years is shown in Table 7.5.

Table 7.5 Sample counts of 'unemployed all year' over time.

	Year			
Unemployment status	1993	1994	1995	1996
Not unemployed (=0)	15451	15373	15406	15406
Unemployed (=1)	280	358	325	325
Total individuals	15731	15731	15731	15731

Note that the binary responses for each of the individuals will be longitudinally correlated. The purpose of the present study is to evaluate the effects of the associated covariates on the unemployment status, by taking the longitudinal correlations of the data into account. As far as the correlation model is concerned, we assume that the data follow one of the three nonstationary, namely AR(1), MA(1), or EQC, correlation models discussed in Section 7.4. As far as the covariates are concerned, we use five important covariates, namely gender, age, geographic location, education level, and marital status of the individual. Although gender, age, and geographic location were held as observed in 1993, education level and marital status are considered to be time-dependent covariates. To shed some light on the nature of the longitudinal relationship between the binary responses 'unemployed all year' and the 5 covariates, we construct appropriate three-way tables for these 5 covariates and the binary response variable 'unemployed all year' for the period from 1993 to 1996. These weighted counts are shown in the appendix, in Tables from 7A to 7E, for the gender, age, region of residence, education level, and marital status, respectively.

Table 7A shows that the proportion of unemployment is more for female than male during all four years from 1993 to 1996. Table 7B shows that there are more 'unemployed all year' individuals in the age group of 25 to 55 which is obvious as this group has the largest range. The proportions of unemployed individuals are, however, also larger for this age group followed by the 16 to 25 age group. The older age group 55 to 65 has the smallest proportions of 'unemployed all year' from 1994 to 1996. The proportions of unemployment appear to decrease over time in all three groups since 1994.

Table 7C shows that the proportion of 'unemployed all year' is the highest in the Atlantic region followed by Quebec, Ontario, Prairie, and BC. Note that this proportion in BC is higher than in Prairies for all four years. Similarly 'unemployed all year' proportions of the Atlantic region are slightly higher than of Quebec except for 1995, Ontario's proportions are being far smaller than those of Quebec as well as the Atlantic region. So, Ontario appears to have a middle place in the country with regard to the 'unemployed all year' status of the individuals. Table 7D helps to understand the effect of education on unemployment over the years. It is clear from this table that the 'high education' group has the smallest 'unemployment all year' rate followed by the 'medium education' group, as expected. These proportions are quite high over the years in the 'lower education' group. Table 7E shows that the proportions of 'unemployed all year' individuals are smaller over the years in the 'married/common law' group, followed by 'widowed', 'single', and 'separated/divorced' groups. More specifically, the proportions are closer to each other between the 'married/common law' and 'widowed' groups, and also between the 'single' and the 'separated/divorced' groups. But when the 'married/common law' or 'widowed' group is compared with the 'single' or 'separated/divorced' group, their proportions appear to be quite different.

7.5.2 Analysis of the SLID Data

For convenience, we rename the five covariates discussed in Section 7.5.1 as follows. First, the gender covariate is represented by x_1 which is 0 for female and 1 for male. For the second covariate, we consider three age groups: group 1 consists of individuals between 16 and 24 inclusive at 1993, group 2 consists of individuals between 25 and 54, and group 3 from 55 to 65. The younger age group 1 is considered to be the reference group. Thus we represent the three groups by x_2 and x_3 so that $x_2 = 0, x_3 = 0$ stands for the individual of the group 1, $x_2 = 1, x_3 = 0$ represents the individual of the group 2, and $x_2 = 0, x_3 = 1$ would identify the individual belonging to the group 3. Similarly, we consider x_4, x_5, x_6, and x_7 to identify an individual from any of the Atlantic, Quebec, Ontario, Prairies, and British Columbia regions. Here we consider the Atlantic region as the reference region with all four variables coded with 0; $x_4 = 1$ and others with 0 will represent the individual from Quebec, and so on. For education level, we have two variables x_8 and x_9 to represent three levels (low, medium and high) of education, lower level being the reference level. Finally, for four marital status: married and common-law spouse, separated and divorced, widow, and single (never married), we use three covariates x_{10}, x_{11}, andx_{12}, respectively, married and common law spouse group being the reference group.

We now compute the effects of these 12 covariates, some being time dependent, on the binary all-year unemployment variable after taking the longitudinal correlations into account. To select one of the three possible nonstationary correlation models, we follow the suggestion from Section 7.4.5 and compute first the initial estimates of the lag correlations $\hat{\rho}_\ell \; \ell(1,\ldots,T-1)$ by (7.91). To be specific, for $\beta = 0$, we compute

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\tilde{y}_{it}\tilde{y}_{i,t+\ell}/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}\tilde{y}_{it}^2/KT},$$

where $\tilde{y}_{it} = [y_{it} - 1/2]/\sqrt{(1/2)(1/2)}$. These values of lag correlations show an exponential decay, a similar pattern for the correlations for the stationary AR(1) model. Note that even though one could use the refined model selection procedure by using (7.92) – (7.94), for simplicity we follow the lead by the pattern found for the initial values for $\hat{\rho}_\ell$ and choose the nonstationary AR(1) model (7.70) to fit the SLID data.

Table 7.6 Nonstationary AR(1) correlation model based estimates of regression and their standard errors, for complete SLID data for the duration from 1993 to 1996.

	Estimate	SE
Male vs Female (x_1)	−0.540	0.072
Age group 2 vs 1 (x_2)	−1.586	0.049
Age group 3 vs 1 (x_3)	−2.168	0.125
Quebec vs Atlantic (x_4)	−0.832	0.083
Ontario vs Atlantic (x_5)	−1.003	0.092
Prairies vs Atlantic (x_6)	−1.854	0.112
BC & Alberta vs Atlantic (x_7)	−1.564	0.159
Education medium vs low (x_8)	−1.604	0.066
Education high vs low (x_9)	−2.454	0.157
Marital status 2 vs 1 (x_{10})	0.206	0.091
Marital status 3 vs 1 (x_{11})	−0.590	0.276
Marital status 4 vs 1 (x_{12})	−0.561	0.095
$\rho_1^{(ns)}$	0.412	–
$\rho_2^{(ns)}$	0.259	–
$\rho_3^{(ns)}$	0.167	–

For the estimation of the regression effects β, we solve the GQL estimating equation (7.85) for β and the moment equation (7.88) for ρ (correlation index parameter) iteratively. Note that to compute the covariance structure in (7.85), we have used the nonstationary AR(1) correlations given by

$$c_{i,u,t}^{(ns)}(x_{iu}, x_{it}, \rho) = \rho^{t-u} \left[\frac{\sigma_{iuu}}{\sigma_{itt}} \right]^{1/2}, \text{ for } u < t,$$

[see (7.73)] with $\sigma_{itt} = \pi_{it}(1 - \pi_{it})$, for example, where

$$\pi_{it} = \frac{\exp(x'_{it}\beta)}{1 + \exp(x'_{it}\beta)}.$$

The estimates of $\beta_1, \ldots, \beta_{12}$, along with their standard errors computed by using (7.87), are shown in Table 7.6. For the sake of completeness, we also show the average lag correlation values computed by

$$\hat{\rho}_\ell^{(ns)} = \hat{\rho}_\ell \frac{1}{K(T-\ell)} \sum_{i=1}^{K} \sum_{u=1}^{T-\ell} \left[\frac{\sigma_{iuu}}{\sigma_{i,u+\ell,u+\ell}} \right]^{1/2},$$

with $\rho_\ell = \rho^\ell$. These estimates are found to be 0.412, 0.259, and 0.167, respectively. Note that the lag correlations, in particular the lag 1 correlation value appears to be large indicating that ignoring the correlation structure, that is, using the independence assumption based approach will produce less efficient regression estimates.

Now, with regard to the interpretation of the GQL regression effects, the negative value −0.540 for the gender effect indicates that the male has lower probability of an all-year unemployment as compared to the female. The negative values -1.586 and -2.168 of β_2 and β_3 indicate that the younger group has higher probability of an all-year unemployment and the probability decreases for older age groups. As far as the effect of geographic location on the all-year unemployment is concerned, it appears that the Prairies had the smallest probability of an all-year unemployment during 1993 to 1996 followed by BC, Ontario, Quebec, and Atlantic provinces. This follows from the fact that the regression estimates for Quebec, Ontario, BC, and Prairies are found to be -0.832, -1.003, -1.854, and -1.564, respectively. The larger negative value -2.454 for β_9 as compared to $\beta_8 = -1.604$ indicates that as the education level gets higher, the probability of an all-year unemployment gets smaller. Finally, with regard to marital status, the positive value 0.206 for β_{10} means that the separated and divorced individuals have higher probability of all-year unemployment as compared to the married and common law spouse group. Similarly, the widowed had less probability of an all-year unemployment as compared to the single but never married individual.

7.6 Application to an Adaptive Clinical Trial Setup

In the last chapter (see Section 6.8), we have discussed the longitudinal count data analysis in an adaptive clinical trial setup, where it is attempted to treat an individual upon arrival with the available better treatment. Note that once a treatment is selected, the individual is treated with the same treatment over the duration of the longitudinal study. In this section, we study a similar problem with an exception that we now collect repeated binary responses as opposed to the repeated count responses. In the cross-sectional setup (i.e., when the individual is treated only once and a binary response is desired), the construction of adaptive design weights for better treatment selection has been discussed by Bandyopadhyay and Biswas (1999), for example. Note that these authors have considered the case where the binary response, on top of treatment effect, is also affected by certain prognostic factors. Sutradhar, Biswas, and Bari (2005) have generalized this idea from the cross-sectional to the longitudinal setup. Here we follow this later work and show (1) how to construct the longitudinal adaptive designs so that a better treatment may be allocated for the incoming individual, and (2) how to estimate the overall treatment effect as well as the effects of the prognostic covariates, by accommodating the longitudinal correlations into account.

7.6.1 Binary Response Based Adaptive Longitudinal Design

Suppose that K independent patients will be treated in the clinical study and T longitudinal binary responses will be collected from each of them. Similar to the count data case, for simplicity, let there be two treatments A and B to treat these patients and A is the better treatment between the two. Next suppose that δ_i refers to the selection of the treatment for the ith $(i = 1, \ldots, K)$ patient, and

$$\delta_i = \begin{cases} 1, & \text{if } i\text{th patient is assigned to } A \\ 0, & \text{if } i\text{th patient is assigned to } B \end{cases}$$

with

$$\Pr(\delta_i = 1) = w_i \text{ and } \Pr(\delta_i = 0) = 1 - w_i. \tag{7.95}$$

Here w_i refers to the better treatment selection probability for the ith patient.

Note that the value of δ_i determines the treatment by which the ith patient will be treated. Now suppose that conditional on δ_i, y_{it} denotes the binary response recorded from the ith patient at time $t (t = 1, \ldots, T)$, and x_{it} denotes the p-dimensional covariate vector corresponding to y_{it}, defined as

$$\begin{aligned} x_{it} &= (\delta_i, x_{it2}, \ldots, x_{itu}, \ldots, x_{itp})' \\ &= (\delta_i, x_{it}^{*'})', \end{aligned} \tag{7.96}$$

where $x_{it}^* = (x_{it2}, \ldots, x_{itu}, \ldots, x_{itp})'$ denote the $p - 1 \times 1$ vector of covariates such as prognostic factors (e.g., age, chronic conditions, and smoking habit) for the ith patient available at time point t. Thus, for $i = 2, \ldots, K$, the distribution of δ_i, that is, the formula of w_i, will depend on $\{\delta_1, \ldots, \delta_{i-1}\}$ and available responses y_{kv} $(k = 1, \ldots, i-1; 1 \leq v \leq T)$ along with their corresponding covariate vector x_{kv}. For $i = 1$, w_1 is assumed to be known.

As far as the availability of the repeated responses is concerned, we assume that for all $i = 1, \ldots, K$, once δ_i becomes known, the repeated binary responses from the ith patient will be available following a binary distribution with conditional mean and variance (conditional on δ_i) given by

$$\begin{aligned} \pi_{it}^*(\delta_i) &= E(Y_{it} | \delta_i, x_{it}^*) = \frac{\exp(\theta_{it})}{1 + \exp(\theta_{it})} \\ \sigma_{itt}^*(\delta_i) &= \text{var}(Y_{it} | \delta_i, x_{it}^*) = \pi_{it}^*(\delta_i)[1 - \pi_{it}^*(\delta_i)], \end{aligned} \tag{7.97}$$

where $\theta_{it} = x_{it}'\beta$, with $x_{it} = (\delta_i, x_{it}^{*'})'$. Also we assume that the pairwise longitudinal correlations between two repeated binary responses are given by

$$\begin{aligned} \text{corr}[(Y_{it}, Y_{iv}) | \delta_i, x_{it}^*, x_{iv}^*] &= \rho_{|t-v|}(\delta_i, x_{it}^*, x_{iv}^*, \rho) \\ &= c_{i,tv}^{(ns)}(\delta_i, x_{it}^*, x_{iv}^*, \rho), \end{aligned} \tag{7.98}$$

where $c_{i,tv}^{(ns)}(\delta_i, x_{it}^*, x_{iv}^*, \rho)$ has the formulas given by (7.73), (7.79), and (7.84) under the nonstationary AR(1), MA(1), and EQC models, respectively. It then follows that the conditional (on δ_i) covariance between y_{it} and y_{iv} is given by

$$\text{cov}[(Y_{it}, Y_{iv}) | \delta_i, x_{it}^*, x_{iv}^*] = c_{i,tv}^{(ns)}(\delta_i, x_{it}^*, x_{iv}^*, \rho)\{\sigma_{itt}^*(\delta_i)\sigma_{ivv}^*(\delta_i)\}^{1/2}, \tag{7.99}$$

with

$$c_{i,tv}^{(ns)}(\delta_i, x_{it}^*, x_{iv}^*, \rho) = \rho^{t-v} \left[\frac{\sigma_{ivv}^*(\delta_i)}{\sigma_{itt}^*(\delta_i)} \right]^{1/2} \quad \text{for } v < t,$$

under the nonstationary binary AR(1) model, for example. Note that Sutradhar, Biswas, and Bari (2005) have used the stationary autocorrelation structure and hence the covariances are approximated as

$$\begin{aligned} \text{cov}[(Y_{it}, Y_{iv})|\delta_i, x_{it}^*, x_{iv}^*] &\simeq c_{i,tv}^*(\rho)\{\sigma_{itt}^*(\delta_i)\sigma_{ivv}^*(\delta_i)\}^{1/2} \\ &= \rho_{|t-v|}\{\sigma_{ivv}^*(\delta_i)\sigma_{itt}^*(\delta_i)\}^{1/2}. \end{aligned} \quad (7.100)$$

7.6.1.1 Simple Longitudinal Play-the-Winner (SLPW) Rule to Formulate w_i

The construction of the SLPW rule in the longitudinal count data setup is described through (a) to (c) in Section 6.8.1.1 and the formulas for w_i are given in Exercise 6.4 for $2 \leq i \leq T$, and in Exercise 6.5 for the case when $T < i \leq K$, K being the number of patients and T is the total number of time points indicating the duration of the longitudinal study. The formulas for w_i in the binary case are similar to those of the Poisson case, except that $I[y_{rt} \leq m_0^*]$ (m_0^* being a threshold number) in the Poisson case is replaced simply by y_{rt} in the binary case, where y_{rt} is the response of the rth individual at time point t. By the same token, $I[y_{rt} > m_0^*]$ from the Poisson case, is now replaced with $1 - y_{rt}$ in the binary case. For convenience, we re-explain the rule in brief and write the formulas for w_i [see also Sutradhar, Biswas, and Bari (2005, Section 2.1)] under the present binary case.

Urn Design for SLPW Rule

1. For the first patient, choose $w_1 = 0.5$ and obtain δ_1 so that $Pr[\delta_1 = 1] = w_1$.
2. Next, for $i = 2, \ldots, K$, the distribution of δ_i will depend on $\{\delta_1, \ldots, \delta_{i-1}\}$ and available responses along with their corresponding covariates. Let this past history be
$$y_{H_{i-1}} \equiv [y_{rt}, x_{rt} \ (r = 1, \ldots, i-1; 1 \leq t \leq \min(T, i-r))].$$
3. As w_i is the probability of selection of the better treatment for the ith patient to be computed based on the history $y_{H_{i-1}}$, it is convenient to compute this w_i by counting two types of balls in an urn [see Wei & Durham (1978), e.g.], the first type being the indicator for the selection of the better treatment A and the second type for the other treatment.
4. The urn will have α balls of each type initially.
5. For a suitable τ value and for available past responses y_{rt}, $y_{rt}\tau$ balls of the same kind by which the rth $(r = 1, \ldots, i-1)$ patient was treated and $(1 - y_{rt})\tau$ balls of the opposite kind are added, at the treatment selection stage for the ith patient.
6. For a suitable quantity u_{rt} [see also Section 7.6.1.2 (b) for its construction] defined such that a larger value of u_{rt} implies the prognostic factor based better condition of the rth $(r = 1, \ldots, i-1)$ past patient, $G - u_{rt}$ balls of the same kind by which the rth patient was treated and u_{rt} balls of the opposite kind are added,

at the treatment selection stage for the ith patient, where $[0,G]$ is the domain of u_{rt}.

Adaptive Design Weights

The above scheme produces the selection probabilities $w_i (i = 1,\ldots,K)$ for the cases $2 \le i \le T$ and $i > T$ as follows.

Case 1. $2 \le i \le T$

Under this case

$$w_i = \Pr(\delta_i = 1|y_{H\,i-1}) = n^*_{i-1,A}(y_{H\,i-1})/n^*_{i-1}, \tag{7.101}$$

where

$$n^*_{i-1} = 2\alpha + \sum_{r=1}^{i-1}\sum_{t=1}^{i-r}(G+\tau) = 2\alpha + (1/2)i(i-1)(G+\tau), \tag{7.102}$$

is the total number of balls in the urn at the selection stage of the ith patient, and

$$n^*_{i-1,A}(y_{H\,i-1}) = \alpha + \sum_{r=1}^{i-1}\sum_{t=1}^{i-r}[\delta_r\{(G-u_{rt}) + y_{rt}\tau\} + (1-\delta_r)\{u_{rt} + (1-y_{rt})\tau\}], \tag{7.103}$$

is the number of balls of the first type that supports the selection of the treatment A.

Case 2. $i > T$

Under this case

$$w_i = \Pr(\delta_i = 1|y_{H\,i-1}) = \tilde{n}_{i-1,A}(y_{H\,i-1})/\tilde{n}_{i-1}, \tag{7.104}$$

where

$$\tilde{n}_{i-1} = 2\alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}(G+\tau) + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}(G+\tau) \tag{7.105}$$

and

$$\begin{aligned}\tilde{n}_{i-1,A}(y_{H\,i-1}) = \alpha &+ \sum_{r=1}^{i-T}\sum_{t=1}^{T}[\delta_r\{(G-u_{rt}) + y_{rt}\tau\} + (1-\delta_r)\{u_{rt} + (1-y_{rt})\tau\}] \\ &+ \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}[\delta_r\{(G-u_{rt}) + y_{rt}\tau\} \\ &+(1-\delta_r)\{u_{rt} + (1-y_{rt})\tau\}],\end{aligned} \tag{7.106}$$

are similar to those of n^*_{i-1} in (7.102) and $n^*_{i-1,A}(y_{Hi-1})$ in (7.103), respectively.

7.6.1.2 Performance of the Adaptive Design

a. Limiting Behavior of Design Weights w_i

Note that it follows from (7.104) that $w_{i+1}/w_i \to 1$ as $i \to \infty$. Again the sequence $\{w_i, i \geq 1\}$ is bounded by 0 from the left and by 1 from the right. Hence there exists a subsequent $w_{k(i)}$ which is convergent. Suppose that it converges to ω. Then from the above limiting result, we have

$$w_{k(i)+1}/w_{k(i)} \to 1$$

as $i \to \infty$, implying for some $\varepsilon > 0$,

$$\omega(1-\varepsilon) \leq \liminf w_{k(i)+1} \leq \limsup w_{k(i)+1} \leq \omega(1+\varepsilon),$$

and hence

$$\limsup w_{k(i)+1} - \liminf w_{k(i)+1} \leq 2\omega\varepsilon.$$

Because ε is arbitrary, we conclude that $\{w_i, i \geq 1\}$ is convergent. Suppose that it converges to ω^*. The formula for ω^* is available from exercise 7.3 [see also Sutradhar, Biswas, and Bari (2005, Section 2.2.1].

b. Allocation Performance (Based on Small Sample) of the Proposed Design: A Simulation Study

In the last subsection, we have computed the limiting value of w_i as $i \to \infty$. As in practice, a large but limited number of patients are considered in a clinical trial study, Sutradhar, Biswas, and Bari (2005) have examined the performance of the proposed adaptive design for $K = 100$ and 200, where K is the total number of patients involved in the clinical trial experiment. We summarize their simulation design and finding as follows.

Simulation Design and Generation of the Design Weights w_i

1. Consider $T = 4$ repeated responses to be collected from each of the K individuals.
2. Consider $p = 4$ covariates; namely 1 treatment covariate (δ_i) and the other 3 prognostic covariates, denoted by x^*_{it2}, x^*_{it3}, and x^*_{it4} for the ith individual at the tth $(t = 1,\ldots,T)$ data collection time.
   a. The values of δ_i for all i $(i = 1,2,\ldots,K)$ are determined based on the adaptive longitudinal design weights

   $$w_i = \Pr(\delta_i = 1|y_{Hi-1}),$$

   constructed in section 7.6.1.1.

b. Suppose that x^*_{it2} represent the chronic disease condition of an incoming patient. Let $c_i \sim \text{bin}(m, \pi)$ (binomial distribution) with $m = 5$ and $\pi = 0.5$. Define

$$x^*_{it2} = \begin{cases} 0 & \text{for } c_i = 0, 1 \\ 1 & \text{for } c_i = 2, \ldots, 5. \end{cases}$$

c. Consider x^*_{it3} and x^*_{it4} to represent an age group of an individual, namely young, middle, and old age groups. let $d_i \sim U[21, 80]$ (uniform distribution). Define

$$(x^*_{it3}, x^*_{it4}) \equiv \begin{cases} (0,0) & \text{for } d_i \in [61, 80] \text{ (old age group)} \\ (0,1) & \text{for } d_i \in [41, 60] \text{ (middle age group)} \\ (1,0) & \text{for } d_i \in [21, 40] \text{ (young age group).} \end{cases}$$

d. In order to compute the adaptive longitudinal design weights w_i [by (7.101) and (7.104)], we also require to define a nonstochastic continuous quantity with domain $[0, G]$, say. More specifically, for a suitable ψ function, we require to construct $u_{rt} = \psi(x^*_{rt2}, x^*_{rt3}, x^*_{rt4})$ that measures the condition of the prognostic covariates x^*_{rt2}, x^*_{rt3}, and x^*_{rt4} so that larger value of u_{rt} implies the better condition of the rth $(r = 1, \ldots, i-1)$ patient. In the simulation study, we choose

$$u_{rt} = (2/(c_r + 1)) + (1/d^*_r),$$

for all t $(1 \leq t \leq \min(T, i - r))$, where c_r is an implicit function of x^*_{rt2}, and similarly d^*_r is an implicit function of x^*_{rt3} and x^*_{rt4}. Note that d^*_i is constructed from d_i as follows

$$d^*_i = \begin{cases} 1 & \text{for } d_i \in [21, 30] \\ 2 & \text{for } d_i \in [31, 40] \\ 3 & \text{for } d_i \in [41, 50] \\ 4 & \text{for } d_i \in [51, 60] \\ 5 & \text{for } d_i \in [61, 70] \\ 6 & \text{for } d_i \in [71, 80]. \end{cases}$$

Further note that as $c_r = 0, 1, \ldots, 5$ and $d^*_r = 1, 2, \ldots, 6$, it then follows that u_{rt} lies in the range of 0 to 3 yielding $G = 3$.

3. Next, for simplicity we consider $\alpha = 1$.
4. Remark that as the limiting value of w_i mainly depends on τ as shown in Exercise 7.3, we consider two values of $\tau = 2$ and 4, one small and the other large.
5. Note that the computation of w_i by (7.101) and (7.104) requires y_{rt}

$$[r = 1, \ldots, i-1; 1 \leq t \leq \min\{T, i-r\}]$$

to be known. For known δ_r $(r = 1, \ldots, i-1)$ the correlated binary responses are generated as follows. First, y_{r1} are generated with probability

$$\Pr(y_{r1} = 1) = \exp(x'_{r.}\beta)/[1 + \exp(x'_{r.}\beta)] = \pi^*_r(\delta_r), \tag{7.107}$$

assuming that $x_{rt} = x_{r.}$ for all $t = 1, \ldots, T$ so that $x_{r.} = (\delta_r, x^*_{r.2}, x^*_{r.3}, x^*_{r.4})'$. Next, we generate $y_{r2}, \ldots, y_{r,\min(T,i-r)}$ following the binary AR(1) model (7.70), for example, with $\pi_{rt} = \pi^*_r(\delta_r)$, for all possible t. In (7.107), use

$$\beta_1 = 1.50, \;\; \beta_2 = 0.0, \;\; \beta_3 = 0.20, \;\; \text{and} \;\; \beta_4 = 0.10.$$

Furthermore, for the ρ parameter in (7.70), we choose the small and large correlation index as $\rho = 0.3$, and 0.7.

Table 7.7 Simulated means and standard errors of δ_s (total number of patients receiving the better treatment) for selected values of the true correlation parameter ρ under AR(1) binary model with $\beta_1 = 1.5$, $\beta_2 = 0.0$, $\beta_3 = 0.2$, and $\beta_4 = 0.1$; and adaptive design parameters $\alpha = 1.0$, $G = 3.0$, and $\tau = 2.0, 4.0$; for different values of $K = 100, 200$

K	τ	ρ	Mean	Standard Error
100	2.0	0.3	58.703	8.505
		0.7	58.632	8.588
	4.0	0.3	62.483	8.779
		0.7	62.348	9.047
200	2.0	0.3	116.660	11.097
		0.7	116.291	11.451
	4.0	0.3	124.693	11.668
		0.7	123.675	12.349

Allocation Performance

Now to examine the allocation performance of the proposed longitudinal adaptive design, we study the distribution of $\delta_s = \sum_{i=1}^{K} \delta_i$ where $w_i = \Pr(\delta_i = 1 | y_{H\,i-1})$ are the design weights defined by (7.101) and (7.104). This we do based on the 1000 simulations. Note that the longitudinal adaptive design proposed in Section 7.6.1.1 is expected to assign more subjects to the better treatment. For this to happen, $\delta_s = \sum_{i=1}^{K} \delta_i$, say, has to be greater than $K/2$.

The values of w_i are calculated following the aforementioned simulation design. Note that once w_i is known, the corresponding δ_i is generated from binary distribution with probability w_i. As mentioned earlier, to understand whether the proposed design can allocate more individuals to the better treatment, we now compute $\delta_s = \sum_{i=1}^{K} \delta_i$ under each of the 1000 simulations. The simulated mean and standard deviation of δ_s for various values of K, τ, and ρ are shown in Table 7.7.

It is clear from Table 7.7 that irrespective of correlation values, the proposed design allocated more individuals to the better treatment A. For example, for $K = 100$, $\tau = 4.0$, and $\rho = 0.7$, 62 individuals out of 100 were assigned to treatment A. Thus relatively more individuals were assigned to the better treatment. Similarly for $K = 200$, $\tau = 4.0$, and $\rho = 0.7$, 124 individuals were allocated to treatment A which is about 62%. Remark that allocation gets better for larger τ. For example, for the same $K = 200$, and $\rho = 0.7$, the allocated number of individuals to treatment A is 116 for the case with $\tau = 2.0$, whereas the allocated number is 124 for $\tau = 4.0$. Thus the proposed design works well in assigning more subjects to a better treatment.

7.6.2 Construction of the Adaptive Design Weights Based Weighted GQL Estimation

Recall from (7.97) that conditional on δ_i, the mean and the variance of Y_{it} are given by

$$E(Y_{it}|\delta_i, x^*_{it}) = \pi^*_{it}(\delta_i) = \frac{\exp(\theta_{it})}{1+\exp(\theta_{it})}, \text{ and } \operatorname{var}(Y_{it}|\delta_i, x^*_{it}) = \pi^*_{it}(\delta_i)[1-\pi^*_{it}(\delta_i)],$$

respectively, with $\theta_{it} = x'_{it}\beta$, where $x_{it} = (\delta_i, x^{*'}_{it})'$. Also, by (7.100), conditional on δ_i, the covariance between y_{iu} and y_{it} has the formula

$$\operatorname{cov}[(Y_{it}, Y_{iv})|\delta_i, x^*_{it}, x^*_{iv}] \simeq \rho_{|t-v|}\{\sigma^*_{ivv}(\delta_i)\sigma^*_{itt}(\delta_i)\}^{1/2}.$$

Note that because in the present adaptive longitudinal setup δ_i depends on

$$\delta_{i-1}, \ldots, \delta_1,$$

finding the unconditional mean and the variance of y_{it} and the unconditional covariance between y_{iu} and y_{it}, will require the unconditional expectation of δ_i to be known, which we compute as follows.

7.6.2.1 Computation of Unconditional Expectation of δ_i : w_{i0}

The distribution of δ_i depends on the past $\delta_{i-1}, \ldots, \delta_1$, thus we write

$$\begin{aligned} w_{i0} = E[\delta_i] &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}(\delta_i) \\ &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_{i-1}|\delta_{i-2},\ldots,\delta_1}[w_i|\delta_{i-1}, \ldots, \delta_1], \end{aligned} \quad (7.108)$$

where w_i has the formula given by (7.101) for $2 \le i \le T$ and by (7.104) for $i > T$. To simplify this expectation, one needs to compute

$$E(\delta_r Y_{rt}) = E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_r|\delta_1,\ldots,\delta_{r-1}} E(\delta_r Y_{rt}|\delta_r, \ldots, \delta_1)$$

$$= E_{\delta_1} E_{\delta_2|\delta_1} \dots E_{\delta_r|\delta_1,\dots,\delta_{r-1}} (\delta_r \pi^*_{rt}(\delta_r)), \text{ for } r = 1,\dots,i-1, \quad (7.109)$$

where

$$\pi^*_{rt} = E(Y_{rt}|\delta_r,\dots,\delta_1) = \exp(x'_{rt}\beta)/(1+\exp(x'_{rt}\beta))$$

with $x_{rt} = (\delta_r, x^*_{rt2}, \dots, x^*_{rtp})'$. Suppose that

$$z_{rt1} = x_{rt}|\delta_r = 1, \text{ and } z_{rt0} = x_{rt}|\delta_r = 0.$$

The expectation in (7.109) then reduces to

$$E(\delta_r Y_{rt}) = w_{r0}\pi_{rt1}, \quad (7.110)$$

where $\pi_{rt1} = \exp(z'_{rt1}\beta)/(1+\exp(z'_{rt1}\beta))$. By similar calculation, it can be shown that

$$E(1-\delta_r)(1-Y_{rt}) = (1-w_{r0})(1-\pi_{rt2}), \quad (7.111)$$

where $\pi_{rt2} = \exp(z'_{rt0}\beta)/(1+\exp(z'_{rt0}\beta))$. Now by applying (7.110) and (7.111) to (7.108), it follows from (7.101) that for $2 \le i \le T$, the unconditional expectation of w_i is given as

$$w_{i0} = \frac{[\alpha + \sum_{r=1}^{i-1}\sum_{t=1}^{i-r}[\{(G-u_{rt}) + \pi_{rt1}\tau\}w_{r0} + \{u_{rt} + (1-p_{rt2})\tau\}(1-w_{r0})]]}{[2\alpha + (1/2)i(i-1)(G+\tau)]}. \quad (7.112)$$

Similarly, it follows from (7.104) that for $i > T$, the unconditional expectation of w_i is given by

$$\begin{aligned} w_{i0} &= \{2\alpha + (G+\tau)T(i-(T+1)/2)\}^{-1} \\ &\times \left[\alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}\{(G-u_{rt}+\pi_{rt1}\tau)w_{r0} + (u_{rt}+(1-\pi_{rt2})\tau)(1-w_{r0})\}\right. \\ &+ \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}\{((G-u_{rt})+\pi_{rt1}\tau)w_{r0} \\ &+ (u_{rt}+(1-\pi_{rt2})\tau)(1-w_{r0})\}]. \end{aligned} \quad (7.113)$$

7.6.2.2 WGQL Estimating Equations for Regression Parameters Including the Treatment Effects

Note that in the conditional mean function $\pi^*_{it}(\delta_i)$ in (7.97), $\beta = [\beta_1, \beta_2, \dots, \beta_p]'$ denotes the effect of $x_{it} = [\delta_i, x^{*'}_{it}]'$ on y_{it}. Here β_1 is the treatment effect and $\beta_2, \dots, \beta_p$, are the effects of $p-1$ prognostic covariates. This is of interest when estimating β after accommodating the longitudinal correlations represented by ρ_ℓ (7.100) for $\ell = 1, \dots, T-1$.

Let $y_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})'$ be a $T \times 1$ vector of repeated binary responses for the ith $(i = 1, \dots, K)$ individual. Note that the ith individual is assigned to treatment

A with probability $w_i = \Pr(\delta i = 1|y_H)$ given by (7.101) for $2 \le i \le T$ and by (7.104) for $i > T$. Here, y_{it} is the tth response of the ith individual. Further note that because w_i depends on the responses from the past $i-1$ patients, the unconditional expectation of y_{it} may be computed as

$$\begin{aligned} E(Y_{it}) &= E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}} E(Y_{it}|\delta_i, \delta_{i-1}, \ldots, \delta_1) \\ &= w_{i0}\pi_{it1} + (1-w_{i0})\pi_{it2} = \bar{\pi}_{it}, \end{aligned} \tag{7.114}$$

where w_{i0} is given by (7.112) for $2 \le i \le T$ and by (7.113) for $i > T$, and π_{it1} and π_{it2} are defined in (7.110) and (7.111), respectively. We now denote by $\bar{\pi}$, the mean vector of y_i. That is,

$$\begin{aligned} \bar{\pi}_i = E(Y_i) &= E(Y_{i1}, \ldots, Y_{iT})' \\ &= [\bar{\pi}_{i1}, \ldots, \bar{\pi}_{it}, \ldots, \bar{\pi}_{iT}]', \end{aligned} \tag{7.115}$$

with $\bar{\pi}_{it}$ as in (7.114) for $t = 1, \ldots, T$.

Next, by using the stationary autocorrelations based conditional autocovariances given by (7.100), one writes the formula for the unconditional covariance between Y_{it} and Y_{iv} as

$$\begin{aligned} \mathrm{cov}(Y_{it}, Y_{iv}) &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}} [\mathrm{cov}(Y_{it}, Y_{iv})|\delta_i, \delta_{i-1}, \ldots, \delta_1)] \\ &\quad + \mathrm{cov}_{\delta_i,\ldots,\delta_1} [E(Y_{it}|\delta_i, \delta_{i-1}, \ldots, \delta_1), E(Y_{iv}|\delta_i, \delta_{i-1}, \ldots, \delta_1)] \\ &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}} [\rho_{|t-v|} \\ &\quad \times \{\pi^*_{it}(\delta_i)(1-\pi^*_{it}(\delta_i))\pi^*_{iv}(\delta_i)(1-\pi^*_{iv}(\delta_i))\}^{1/2}] \\ &\quad + cov_{\delta_i,\ldots,\delta_1} [\pi^*_{it}(\delta_i), \pi^*_{iv}(\delta_i)], \end{aligned} \tag{7.116}$$

where by (7.97) we have used $E(Y_{it}|\delta_i, \ldots, \delta_1) = \pi^*_{it}(\delta_i) = \exp(x'_{it}\beta)/(1+\exp(x'_{it}\beta))$ and $\mathrm{var}(Y_{it}|\delta_i, \ldots, \delta_1) = \pi^*_{it}(\delta_i)(1-\pi^*_{it}(\delta_i))$. After some algebra, by (7.100), the equation (7.116) reduces to

$$\begin{aligned} \mathrm{cov}(Y_{it}, Y_{iv}) &= \rho_{|t-v|} \Big[w_{i0}\{\pi_{it1}(1-\pi_{it1})\pi_{iv1}(1-\pi_{iv1})\}^{1/2} \\ &\qquad + (1-w_{i0})\{\pi_{it2}(1-\pi_{it2})\pi_{iv2}(1-\pi_{iv2})\}^{1/2} \Big] \\ &\qquad + w_{i0}\{\pi_{it1}\pi_{iv1}\} + (1-w_{i0})\{\pi_{it2}\pi_{iv2}\} - \bar{\pi}_{it}\bar{\pi}_{iv} \\ &= \bar{\sigma}_{itv}(w_{i0}). \end{aligned} \tag{7.117}$$

When $t = v$, the covariance $\bar{\sigma}_{itv}(w_{i0})$ in (7.117) reduces to the variance of y_{it} given by

$$\mathrm{var}(Y_{it}) = \bar{\sigma}_{itt}(w_{i0}) = \bar{\pi}_{it}(1-\bar{\pi}_{it}). \tag{7.118}$$

Let $\bar{\Sigma}_i(w_{i0}, \rho)$ denote the covariance matrix of y_i, which may be expressed as

$$\bar{\Sigma}_i(w_{i0},\rho) = \mathrm{cov}(Y_i) = (\bar{\sigma}_{itv}(w_{i0})),$$

for $t,v = 1,\ldots,T$, where $\bar{\sigma}_{itv}(w_{i0})$ are given by (7.117) and (7.118).

Next for known $\bar{\Sigma}_i(w_{i0},\hat{\rho})$, we write the generalized quasi-likelihood (GQL) estimating equation for β as

$$\sum_{i=1}^{K}(\partial\bar{\pi}_i^{'}(w_{i0})/\partial\beta)\bar{\Sigma}_i^{-1}(w_{i0},\hat{\rho})\,(y_i-\bar{\pi}_i(w_{i0})) = 0 \qquad (7.119)$$

[Sutradhar (2003)], where $\bar{\pi}_i(w_{i0})$ is the $T\times 1$ vector given by (7.115) and $\partial\bar{\pi}_i^{'}(w_{i0})/\partial\beta$ is the $p\times T$ first derivative vector of $\bar{\pi}_i^{'}(w_{i0})$ with respect to β. Note that to be precise, we refer to (7.119) as the weighted GQL estimating equation. This is for the fact that the binary probabilities in (7.119) are adaptive design weights dependent.

Now to solve (7.119) for β, one may consider the following three scenarios: first, for some initial β, w_{i0} is known in the spirit of GEE; second, w_{i0} is unknown but it can be replaced with adaptive design weight w_i as $E(w_i) = w_{i0}$; third, w_{i0} is an unknown function of β. Here we use the second option and refer to Sutradhar, Biswas, and Bari (2005) for details on all three scenarios. Suppose that $\hat{\beta}_{WGQL}$ denotes the solution of (7.119) that may be obtained by using iterative equation

$$\hat{\beta}_{(m+1)_{GQL}} = \hat{\beta}_{(m)_{GQL}} + \left[\sum_{i=1}^{K}(\partial\bar{\pi}_i^{'}(w_{i0})/\partial\beta)\bar{\Sigma}_i^{-1}(w_{i0},\hat{\rho})(\partial\bar{\pi}_i(w_{i0})/\partial\beta^{'})\right]_m^{-1}$$
$$\times\left[\sum_{i=1}^{K}(\partial\bar{\pi}_i^{'}(w_{i0})/\partial\beta)\bar{\Sigma}_i^{-1}(w_{i0},\hat{\rho})\,(y_i-\bar{\pi}_i(w_{i0}))\right]_m, \qquad (7.120)$$

where $\hat{\beta}_{(m)_{GQL}}$ is the value of β at the mth iteration and $[\cdot]_m$ denotes that the expression within brackets is evaluated at $\hat{\beta}_{(m)_{GQL}}$. By (7.114), the first derivative in (7.120) has the formula

$$\begin{aligned}\partial\bar{\pi}_i^{'}(w_{i0})/\partial\beta &= \partial\bar{\pi}_i^{'}(w_{i0})/\partial\beta\,|_{w_{i0}=w_i}\\ &= w_i(\partial\pi_{i\cdot 1}^{'}/\partial\beta) + (1-w_i)(\partial\pi_{i\cdot 2}^{'}/\partial\beta),\end{aligned} \qquad (7.121)$$

where

$$\pi_{i\cdot 1} = (\pi_{i11},\ldots,\pi_{it1},\ldots,\pi_{iT1})^{'}, \text{ and } \pi_{i\cdot 2} = (\pi_{i12},\ldots,\pi_{it2},\ldots,\pi_{iT2})^{'},$$

with

$$\pi_{it1} = \exp(z_{it1}^{'}\beta)/(1+\exp(z_{it1}^{'}\beta)), \text{ and } \pi_{it2} = \exp(z_{it0}^{'}\beta)/(1+\exp(z_{it0}^{'}\beta)),$$

where $z_{it1} = x_{it}|_{\delta_i=1}$ and $z_{it0} = x_{it}|_{\delta_i=0}$. It then follows from (7.121) that

$$\partial \bar{\pi}_i^{'}(w_{i0})/\partial \beta = w_i Z_i' A_{i1} + (1-w_i) Z_i^{*'} A_{i2} = C_i, \tag{7.122}$$

where

$$Z_i' = (z_{i11}, \ldots, z_{it1}, \ldots, z_{iT1}), \quad \text{and } Z_i^{*'} = (z_{i10}, \ldots, z_{it0}, \ldots, z_{iT0}),$$

are $p \times T$ matrices, and

$$A_{i1} = \text{diag}[\pi_{i11}(1-\pi_{i11}), \ldots, \pi_{iT1}(1-\pi_{iT1})],$$

and

$$A_{i2} = \text{diag}[\pi_{i12}(1-\pi_{i12}), \ldots, \pi_{iT2}(1-\pi_{iT2})],$$

are $T \times T$ matrices.

Note that solving the iterative equation (7.120) for β requires the knowledge of $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_\ell, \ldots, \hat{\rho}_{T-1})$ where $\hat{\rho}_\ell$ $(\ell = 1, \ldots, T-1)$ may be obtained consistently as in the next section, by using the so-called method of moments.

7.6.2.2.1 Moment Estimates for Longitudinal Correlations

For a given value of the estimate of β, we now obtain a moment estimator $\hat{\rho}$, which is consistent for ρ. To be specific, by (7.117) we write the moment estimator as

$$\hat{\rho}_\ell = [a_\ell/s - b_\ell/\xi]/[c_\ell/\xi], \tag{7.123}$$

where

$$\begin{aligned}
a_\ell &= \sum_{i=1}^{K} \sum_{|t-v|=\ell} [(y_{it} - \bar{\pi}_{it})(y_{iv} - \bar{\pi}_{iv})]/K(T-\ell) \\
s &= \sum_{i=1}^{K} \sum_{t=1}^{T} [y_{it} - \bar{\pi}_{it}]^2/KT \\
b_\ell &= \sum_{i=1}^{K} \sum_{|t-v|=\ell} [w_{i0}\bar{\pi}_{it1}\bar{\pi}_{iv1} + (1-w_{i0})\bar{\pi}_{it2}\bar{\pi}_{iv2} - \bar{\pi}_{it}\bar{\pi}_{iv}]/K(T-\ell) \\
\xi &= \sum_{i=1}^{K} \sum_{t=1}^{T} [\bar{\pi}_{it}(1-\bar{\pi}_{it})]/KT,
\end{aligned}$$

and

$$\begin{aligned}
c_\ell = \sum_{i=1}^{K} \sum_{|t-v|=\ell} \Big[& w_{i0}\{\pi_{it1}(1-\pi_{it1})\pi_{iv1}(1-\pi_{iv1})\}^{1/2} \\
& + (1-w_{i0})\{\pi_{it2}(1-\pi_{it2})\pi_{iv2}(1-\pi_{iv2})\}^{1/2} \Big] /K(T-\ell).
\end{aligned}$$

Note that the w_{i0} in b_ℓ and c_ℓ may be replaced with data based adaptive design weight w_i $(i = 1,\ldots,K)$.

7.6.2.2.2 Asymptotic Variances of the WGQL Regression Estimates

By using the multivariate central limit theorem [see Mardia, Kent and Bibby (1979, p. 51)], it may be shown that for large K, $\hat{\beta}_{WGQL}$ obtained from (7.120) have asymptotically p-dimensional normal distribution with mean β and $p \times p$ covariance matrix V which can be estimated as

$$\hat{V} = \hat{\text{var}}(\hat{\beta}_{WGQL}) = \left[\sum_{i=1}^{K} C_i \bar{\Sigma}_i^{-1}(w_i,\hat{\rho})C_i^{'}\right]^{-1}, \tag{7.124}$$

where $\bar{\Sigma}_i(w_i,\hat{\rho})$ is obtained from $\bar{\Sigma}_i(w_{i0},\hat{\rho})$ by replacing w_{i0} with its data based estimate w_i, and C_i is given by (7.122).

We remark here that it has been demonstrated by Sutradhar, Biswas, and Bari (2005) through a simulation study that the WGQL approach performs very well in estimating the treatment as well as other regression effects. See their Tables 2 and 3 for details. It is also demonstrated by these authors (see their Table 4) that ignoring adaptive design weights w_i, that is, using random design weight $w_i = 0.5$ causes mean squared efficiency loss in treatment and other regression effects estimation.

7.7 More Nonstationary Binary Correlation Models

In some longitudinal studies for binary data, the expectation of the binary response variable of an individual at a given point of time may depend on the covariate history up to the present time. By the same token, the variance at a given point of time and the correlation of the two responses at two given time points may also depend on the history of the time-dependent covariates of the individual. In this section, we discuss two such binary dynamic models, one linear and the other nonlinear, by nature. The linear binary dynamic regression (LBDR) model and its basic properties along with inferences for the regression effects are discussed in Section 7.7.1, whereas details on a nonlinear BDR (NLBDR) model are given in Section 7.7.2.

7.7.1 Linear Binary Dynamic Regression (LBDR) Model

Let $\{y_{it}, t = 1, \ldots, T\}$ be a sequence of repeated binary responses and $x_{it} = (x_{it1},\ldots,x_{tp})^{'}$ be the p-dimensional vector of covariates corresponding to y_{it}. Also let $\beta = (\beta_1,\ldots,\beta_p)^{'}$ be the p-dimensional effect of the covariates x_{it} on y_{it}, for every individual $i = 1,\ldots,K$. Suppose that for the ith individual, there exists a binary series $\{\varepsilon_{i,2t}, t = 1, \ldots, T\}$ such that

$$\begin{aligned} Pr[\varepsilon_{i,2t} = 1] &= \pi_{it} \\ &= \exp(x_{it}' \beta)/[1 + \exp(x_{it}' \beta)], \end{aligned} \tag{7.125}$$

that is, $\varepsilon_{i,2t} \sim b(\pi_{it})$. Further suppose that $y_{i1} \sim b(\pi_{i1})$. One may now write a LBDR model for $\{y_{it}\}$ as a linear mixture of $y_{i,t-1}$ and $\varepsilon_{i,2t}$ defined as

$$y_{it} = \varepsilon_{i,1t} y_{i,t-1} + (1 - \varepsilon_{i,1t})\varepsilon_{i,2t}, \tag{7.126}$$

where for a suitable $0 < \rho < 1$, $\varepsilon_{i,1t}$ denotes a binary variable with mixture probability ρ, that is, $\varepsilon_{i,1t} \sim b(\rho)$. Further note that $\varepsilon_{i,1t}$ and $\varepsilon_{i,2t}$ are independent. It then follows that $\{y_{it}\}$ generated by (7.126) constitute a sequence of repeated binary observations with nonstationary marginal mean and variance given by

$$E[Y_{it}] = \mu_{it} = \sum_{j=1}^{t-1} (\pi_{ij} - \pi_{i,j+1})\rho^{t-j} + \pi_{it} \tag{7.127}$$

$$\text{var}[Y_{it}] = \sigma_{itt} = \mu_{it}(1 - \mu_{it}). \tag{7.128}$$

Note that as opposed to the longitudinal setup, this linear dynamic mixture model (7.126) has been discussed by Tong [1990, model (4), Table 3.1, p.113], among others, in the time series setup. See also Tagore and Sutradhar (2009) for inferences in correlated binary regression model in time series setup.

7.7.1.1 Autocorrelation Structure

It follows that the LBDR model (7.126) yields the lag ℓ ($\ell = 1 \ldots, T-1$) autocorrelation between y_{it} and $y_{i,t-\ell}$ $(t = 2, \ldots, T)$ given by

$$\text{corr}[Y_{it}, Y_{i,t-\ell}] = \rho_\ell(y) = \frac{\mu_{i,t-\ell}\{\rho^\ell + (1-\rho)\sum_{j=0}^{l-1} \rho^j \, \pi_{i,t-j} - \mu_{it}\}}{[\sigma_{itt} \; \sigma_{i,t-\ell,t-\ell}]^{\frac{1}{2}}}, \tag{7.129}$$

where μ_{it} and σ_{itt} are given by (7.127) and (7.128), respectively.

Note that under the present dynamic model (7.126), the mean of y_{it} is π_{it} plus a weighted sum of successive differences of π_{ij} and $\pi_{i,j+1}$ for $j = 1, \ldots, t-1$, where weights follow an exponential function in mixture probability ρ. Thus the linear mixture model (7.126) has the mean at a given time t which depends on the past means, that is, on the past history. By the same token, the variance given by (7.127) also depends on the past history. Also, it is clear that this dynamic model in (7.126) is different from the nonstationary conditionally linear binary models discussed in Section 7.4. It was demonstrated in Section 7.4 that these later models unlike (7.126) produce means and variances at a given time t that depend on the covariates collected at the same time point t only. The nonlinear dynamic model discussed in Section 7.8 has properties similar to the model (7.126). Further note that for $\rho \longrightarrow 0$, the mean of y_{it} under the model (7.126), however, tends to π_{it}. That is, in such a case the past binary contributes very little. For $\rho \longrightarrow 1$, the mean μ_{it} tends to π_{i1}. That is, the series depends mostly on the initial binary response.

As far as the autocorrelations given by (7.129) of the repeated binary responses for the ith individual are concerned, it may be shown that $\rho_\ell(y)$ satisfies a narrower range than -1 to 1. For example, for a given t,

$$\rho_\ell(y) \longrightarrow 0 \quad \text{as } \rho \longrightarrow 0 \tag{7.130}$$

and

$$\rho_\ell(y) \longrightarrow 1 \quad \text{as } \rho \longrightarrow 1. \tag{7.131}$$

Thus for all t and any $0 < \rho < 1$, the lag correlation $\rho_\ell(y)$ has the range between 0 and 1, for a given individual.

7.7.1.2 GQL and Conditional GQL (CGQL) Approaches for Parameter Estimation

GQL Estimation for β

Note that it follows from (7.127) that the response vector $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ has the mean $\mu_i = (\mu_{i1}, \ldots, \mu_{it}, \ldots, \mu_{iT})'$. Furthermore, let Σ_i denote the $T \times T$ covariance matrix of y_i. To be specific, the diagonal elements of this matrix are given by (7.128) for all $t = 1, \ldots, T$, for a given individual $i = 1, \ldots, K$. For $u < t$, the off-diagonal elements of Σ_i are given by

$$\begin{aligned}\sigma_{i,t-u,t} = \sigma_{i,t,t-u} &= \text{cov}(Y_{i,t-u}, Y_{it}) \\ &= \mu_{i,t-u}\,[\rho^u + (1-\rho)\sum_{j=0}^{u-1} \rho^j \pi_{i,t-j} - \mu_{it}],\end{aligned} \tag{7.132}$$

[see also (7.129)]. We may then exploit a two-moments based GQL approach to estimate β. More specifically, the GQL estimating equation for β is written as

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1} (y_i - \mu_i) = 0, \tag{7.133}$$

where for $t = 2, \ldots, T$,

$$\frac{\partial \mu_{it}}{\partial \beta} = \pi_{it}(1-\pi_{it})x_{it}' + \sum_{j=1}^{t-1} [\pi_{ij}(1-\pi_{ij})x_{ij}' - \pi_{i,j+1}(1-\pi_{i,j+1})x_{i,j+1}']\rho^{t-j}$$

whereas

$$\frac{\partial \mu_{i1}}{\partial \beta} = \pi_{i1}(1-\pi_{i1})x_{i1}'.$$

Note that the solution of (7.133) produces a consistent as well as a highly efficient estimator for β as compared to moment estimator, for example. This is because the GQL estimating equation (7.133), similar to that of (7.85), is unbiased for zero and it uses the inverse of the covariance matrix as a weight function in the estimating equation.

CGQL Estimation for β

In the time series setup, the GQL approach, however, may encounter computational difficulties when the Σ_1 (as $i = 1$ only) matrix has large dimension, that is, when the time series is long. As a remedy, Tagore and Sutradhar (2009, Section 3.1.2, p. 888) have used a conditional GQL (CGQL) approach for the estimation of β. One may follow this approach and write the CGQL estimating equation for β in the present longitudinal setup, as

$$\sum_{i=1}^{K} \frac{\partial \mu'_{i(c)}}{\partial \beta} \Sigma_{i(c)}{}^{-1}(y_i - \mu_{i(c)}) = 0, \tag{7.134}$$

where $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ is the vector of observations as before, $\mu_{i(c)}$ is the conditional mean of y_i; that is,

$$\begin{aligned} \mu_{i(c)} &= E[Y_{i1}, Y_{i2}|Y_{i1}, \ldots, Y_{it}|Y_{i,t-1}, \ldots, Y_{iT}|Y_{i,T-1}]' \\ &= [\lambda^*_{i1}, \lambda^*_{i2}, \ldots, \lambda^*_{it}, \ldots, \lambda^*_{iT}]', \end{aligned} \tag{7.135}$$

where by the model (7.125) – (7.126)

$$\lambda^*_{it} = \begin{cases} Pr[y_{i1} = 1] = \pi_{i1}, & \text{for } t = 1 \\ Pr[y_{it} = 1|y_{i,t-1}] = \pi_{it} + \rho(y_{i,t-1} - \pi_{it}), & \text{for } t = 2, \ldots, T, \end{cases} \tag{7.136}$$

are the same as in (7.70), with $\pi_{it} = \exp(x'_{it}\beta)/[1 + \exp(x'_{it}\beta)]$ for all $t = 1, 2, \ldots, T$,

In (7.134), unlike under (7.70), $\Sigma_{i(c)}$ denotes the covariance matrix of the elements of y_i conditional on the past history. To be specific, $\Sigma_{i(c)}$ has a diagonal form given by

$$\begin{aligned} \Sigma_{i(c)} &= \mathrm{cov}[Y_{i1}, Y_{i2}|Y_{i1}, \ldots, Y_{it}|Y_{i,t-1}, \ldots, Y_{iT}|Y_{i,T-1}]' \\ &= \mathrm{diag}[\mathrm{var}(Y_{i1}), \mathrm{var}(Y_{i2}|Y_{i1}), \ldots, \mathrm{var}(Y_{it}|Y_{i,t-1}), \ldots, \mathrm{var}(Y_{iT}|Y_{i,T-1})] \\ &= \mathrm{diag}[\sigma_{i,11(c)}, \sigma_{i,22(c)}, \ldots, \sigma_{i,tt(c)}, \ldots, \sigma_{i,TT(c)}]' \end{aligned} \tag{7.137}$$

with $\sigma_{i,tt(c)} = \lambda^*{}_{it}(1 - \lambda^*{}_{it})$ for all $t = 1, 2, \ldots, T$. Furthermore, in (7.134)

$$\frac{\partial \mu'_{i(c)}}{\partial \beta} = [\frac{\partial \mu_{i(c)}(1)}{\partial \beta}, \ldots, \ldots, \frac{\partial \mu_{i(c)}(t)}{\partial \beta}, \ldots, \frac{\partial \mu_{i(c)}(T)}{\partial \beta}]', \tag{7.138}$$

where

$$\frac{\partial \mu_{i(c)}(t)}{\partial \beta} = (1 - \rho)\frac{\partial \pi_{it}}{\partial \beta}$$

by (7.135), with $\partial\pi_{it}/\partial\beta = \pi_{it}(1-\pi_{it})x_{it}$.

The CGQL estimating equation (7.134) can be solved iteratively. Let $\hat{\beta}_{CGQL}$ be the solution, which may be obtained by using the iterative equation

$$\hat{\beta}_{CGQL}(r+1) = \hat{\beta}_{CGQL}(r) + \left[\sum_{i=1}^{K}\left(\frac{\partial\mu'_{i(c)}}{\partial\beta}\Sigma_{i(c)}^{-1}\frac{\partial\mu_{i(c)}}{\partial\beta}\right)^{-1}\sum_{i=1}^{K}\frac{\partial\mu'_{i(c)}}{\partial\beta}\Sigma_{i(c)}{}^{-1}(y_i-\mu_{i(c)})\right]_{\beta=\hat{\beta}_{CGQL(r)}} \quad (7.139)$$

Note that because $E[Y_{it}|Y_{i,t-1}-\mu_{i(c)}(t)] = 0$, the CGQL estimating equation is unbiased and it produces a consistent estimator of β. Further note that because $\Sigma_{i(c)}$ is a diagonal matrix, the solution of the CGQL estimating equation (7.134) for β is straight forward.

Moment Estimating Equation for ρ

Note that the estimation of β by using either the GQL estimating equation (7.133) or the CGQL estimating equation (7.134) requires the mixture probability ρ to be known. In practice, however, this ρ is unknown.

Further note that when β is estimated by using the GQL estimating equation (7.133), it is reasonable to estimate the ρ parameter by solving a moment equation in a lag 1 sample correlation given by

$$\hat{\rho}_1 - \rho_1 = 0, \quad (7.140)$$

where for given β, ρ_ℓ from (7.129) is a function of ρ. But this estimate does not have closed-form expression because the unconditional mean given in (7.127) is a polynomial function in ρ, and also the lag correlation ρ_1 involves the unconditional mean in a complicated way. For practical convenience, an approximate estimate of ρ is found in Exercise 7.4. Note that the estimate of ρ found from exercise 7.4 is then used in (7.133) to obtain an improved estimate of β, which in turn is used in exercise 7.4 to improve the estimate for ρ. This constitutes a cycle of iteration, and the cycle continues until the convergence of the estimate is achieved.

In the CGQL approach, one may, however, easily compute a moment estimate of ρ by minimizing the conditional mean squared error $\sum_{i=1}^{K}\sum_{t=2}^{T}(y_{it}-\lambda_{it}^{*})^2$. The estimating formula for ρ is given by

$$\hat{\rho} = \frac{\sum_{i=1}^{K}\sum_{t=2}^{T}(y_{it}-\pi_{it})(y_{i,t-1}-\pi_{it})}{\sum_{i=1}^{K}\sum_{t=2}^{T}(y_{i,t-1}-\pi_{it})^2}. \quad (7.141)$$

This estimate of ρ is then used in (7.134) to obtain an improved estimate of β, which in turn is used in (7.141) to improve the estimate for ρ. This constitutes a cycle of iteration. The cycle continues until the convergence of the estimate is achieved.

7.7.2 A Binary Dynamic Logit (BDL) Model

As opposed to the LBD model (7.126), there exists a more flexible correlation structure based nonlinear binary dynamic model. In a time series setup, this type of nonlinear dynamic model was studied by many econometricians. See, for example, Amemiya (1985, p. 422), Manski (1987), and Farrell and Sutradhar (2006), among others. For discussion on this type of nonlinear binary dynamic model in the longitudinal set up, we refer to Sutradhar and Farrell (2007), for example. This model may be written as

$$\begin{aligned} p_{i1} &= Pr[y_{i1} = 1] = \pi_{i1} = \exp(x'_{i1}\beta)/[1+\exp(x'_{i1}\beta)] \\ p_{it|t-1} &= Pr[y_{it} = 1|y_{i,t-1}] = \frac{\exp(x'_{it}\beta + \theta y_{i,t-1})}{1+\exp(x'_{it}\beta + \theta y_{i,t-1})}, \end{aligned} \tag{7.142}$$

for $t = 2,\ldots,T$, where θ may be referred to as the dynamic dependence parameter. Note that this lag 1 dependence model (7.142) is a special case of a full lag dependence model defined as

$$\begin{aligned} p_{it|t-1,t-2,\ldots,1} &= Pr[y_{it} = 1|y_{i,t-1},\ldots,y_{i1}] \\ &= \frac{\exp(x'_{it}\beta + \theta_1 y_{i,t-1} + \theta_2 y_{i,t-2} + \ldots + \theta_{t-1} y_{i1})}{1+\exp(x'_{it}\beta + \theta y_{i,t-1} + \theta_2 y_{i,t-2} + \ldots + \theta_{t-1} y_{i1})}, \end{aligned} \tag{7.143}$$

which is a nonlinear probability function, whereas the model in (7.39) considered by Qaqish (2003) is linear by nature. Note that as opposed to the conditional linear probability model (7.39), this model in (7.143) is valid for any range for the dynamic dependence parameters $\theta_1,\ldots,\theta_{T-1}$. Consequently, the correlations computed from the nonlinear logistic model (7.143) must satisfy the range from -1 to $+1$. For simplicity, here we deal with the lag 1 dependence model (7.142) and provide its basic properties as follows [see also Sutradhar and Farrell (2007)].

7.7.2.1 Basic Properties of the Lag 1 Dependence Model (7.142)

Unconditional Mean and Variance

For p_{i1} and $p_{it|t-1}$ defined as in (7.142), let

$$\tilde{p}_{it} = p_{it|t-1}\big|_{y_{i,t-1}=1} = \frac{\exp(x'_{it}\beta + \theta)}{1+\exp(x'_{it}\beta + \theta)}. \tag{7.144}$$

It then follows that the unconditional mean of y_{it} satisfies the recursive relationship

$$\mu_{it} = E(Y_{it}) = Pr(y_{it} = 1) = \pi_{it} + \mu_{i,t-1}(\tilde{p}_{it} - \pi_{it}), \text{ for } t = 2,\ldots,T, \tag{7.145}$$

with $\mu_{i1} = \pi_{i1}$, where

$$\pi_{it} = \frac{\exp(x'_{i1}\,\beta)}{[1+\exp(x'_{i1}\,\beta)]},$$

for all $t = 1, \ldots, T$. Note that the expectation in (7.145) may be derived by using the conditioning argument. For example,

$$\begin{aligned} E[Y_{i2}] &= E_{Y_{i1}} E[Y_{i2}|y_{i1}] \\ &= E_{Y_{i1}}[p_{i2|1}(y_{i1})], \\ &= \sum_{y_{i1}=0}^{1} [p_{i2|1}(y_{i1})] \pi_{i1}{}^{y_{i1}} (1-\pi_{i1})^{1-y_{i1}}, \end{aligned} \tag{7.146}$$

where by the model (7.142), $p_{i2|1}(y_{i1})$ as a function of y_{i1} is given by

$$p_{i2|1}(y_{i1}) = \frac{\exp(x'_{it}\beta + \theta y_{i1})}{1+\exp(x'_{it}\beta + \theta y_{i1})}.$$

Furthermore, because

$$p_{i2|1}(1) = \tilde{p}_{i2} \text{ and } p_{i2|1}(0) = \pi_{i2},$$

it then follows from (7.146) that

$$\begin{aligned} E[Y_{i2}] &= \tilde{p}_{i2}\pi_{i1} + \pi_{i2}(1-\pi_{i1}) \\ &= \pi_{i2} + \pi_{i1}[\tilde{p}_{i2} - \pi_{i2}], \end{aligned} \tag{7.147}$$

yielding

$$\mu_{i2} = \pi_{i2} + \mu_{i1}[\tilde{p}_{i2} - \pi_{i2}].$$

By similar arguments, the unconditional expectation of y_{it}, that is,

$$E[Y_{it}] = E_{Y_{i1}} E_{Y_{i2}|y_{i1}} \ldots E_{Y_{it}|y_{i,t-1}}[Y_{it}|y_{i,t-1}]$$

can be derived in the form (7.145). The variance of y_{it} has the formula

$$\sigma_{itt} = \text{var}[Y_{it}] = \mu_{it}[1-\mu_{it}], \tag{7.148}$$

where μ_{it} is the unconditional expectation given by (7.145).

Covariances and Correlations

For $u < t$, by computing the expectation of the product of y_{iu} and y_{it} following a conditional argument, that is,

$$E[Y_{iu}Y_{it}] = E_{Y_{iu}}\left[Y_{iu}E_{Y_{i,u+1}|y_{iu}} \ldots E_{Y_{it}|y_{i,t-1}}\{Y_{it}|y_{i,t-1}, \ldots, y_{iu}\}\right],$$

the covariance between y_{iu} and y_{it} under the model (7.142) is obtained as

$$\text{cov}(Y_{iu}, Y_{it}) = E[Y_{iu}Y_{it}] - \mu_{iu}\mu_{it} = \sigma_{iut} = \mu_{iu}(1-\mu_{iu}) \prod_{j=u+1}^{t} (\tilde{p}_{ij} - \pi_{ij}), \quad (7.149)$$

where μ_{iu} is given by (7.145), and $\tilde{p}_{ij}$ and π_{ij}, respectively, have the formulas

$$\tilde{p}_{ij} = \frac{\exp(x'_{ij}\beta + \theta)}{1 + \exp(x'_{ij}\beta + \theta)} \text{ and } \pi_{ij} = \frac{\exp(x'_{ij}\beta)}{1 + \exp(x'_{ij}\beta)}.$$

Consequently, one obtains the lag $(t-u)$ autocorrelation between y_{iu} and y_{it} as

$$\text{corr}(Y_{iu}, Y_{it}) = \sqrt{\frac{\mu_{iu}(1-\mu_{iu})}{\mu_{it}(1-\mu_{it})}} \prod_{j=u+1}^{t} (\tilde{p}_{ij} - \pi_{ij}), \quad (7.150)$$

which satisfies the full range from -1 to 1, as $0 < \tilde{p}_{ij}, \pi_{ij} < 1$ [see also Sutradhar and Farrell (2007)].

Note that the nonlinear BDL model (7.142) is more appropriate for situations where the mean and the variance at a given point of time are thought to be influenced by the past means and variances. This is technically evident from the formulas for the marginal means and variances shown in (7.145) and (7.148), respectively. In practice, one encounters this situation, for example, in socioeconomic studies involving growth in gross domestic products (GDP), where such growth at a given year is most likely to be influenced by the GDP growth over the past. Similarly, in a biomedical such as asthma study, the mean asthma status of a patient at a given week is most likely to be influenced by the average asthma status of the individual in the past. Further note that the autocorrelation structure (7.150) of the model is quite flexible. According to this model, one does not need to know whether correlations follow any known Gaussian type such as AR(1), MA(1), and EQC models. Moreover, unlike the nonstationary binary correlation models discussed in Section 7.4, the BDL model (7.142) accommodates correlations with full range from -1 to $+1$ as shown in (7.150). It has been demonstrated by Farrell and Sutradhar (2006, Table 2) that the correlations generated by the model (7.142) may lie outside the ranges of correlations produced by the linear dynamic conditional probability model (7.139) [see also Qaqish (2003)]. This shows that the BDL model is more appealing to use in practice as opposed to the conditional linear dynamic models discussed in Section 7.4.

7.7.2.2 Estimation of the Parameters of the BDL Model

To fit the BDL model to the longitudinal binary data, it is necessary to estimate the regression effects β and the dynamic dependence parameter θ, consistently and efficiently. This we do in the next two sections by using a generalized quasi-likelihood estimation approach. A standard simple GQL as well as a so-called optimal GQL estimating equations is considered. We also consider the maximum likelihood estimation. It is demonstrated that the OGQL estimates are the same as ML estimates. Note that even though the ML estimation is quite manageable for the BDL fixed effects model (7.142), this ML approach may be complicated under the BDL mixed effects model. This we discuss in Chapter 9.

7.7.2.2.1 GQL Estimation

Let $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ denote the vector of T repeated binary responses with y_{it} for $t = 2, \ldots, T$, following the nonlinear dynamic model (7.142). Also, let $\mu_i = (\mu_{i1}, \ldots, \mu_{it}, \ldots, \mu_{iT})'$ be the unconditional mean of the response vector y_i, where μ_{it} is the expectation of y_{it}, which is computed by (7.145). That is,

$$\mu_{it} = \pi_{it} + \mu_{i,t-1}(\tilde{p}_{it} - \pi_{it}).$$

Furthermore, let $\Sigma_i = (\sigma_{iut})$ be the $T \times T$ covariance matrix of y_i, where σ_{itu} for $u < t$ is defined by (7.149). As far as the diagonal elements of the Σ_i matrix are concerned , they are the variances of the repeated data, and they are given in (7.148).

The generalized quasi-likelihood estimate of $\zeta = (\beta', \theta)'$ is now obtained by solving the estimating equation

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \zeta} \Sigma_i^{-1} (y_i - \mu_i) = 0, \tag{7.151}$$

[Sutradhar (2003)] where $\partial \mu_i' / \partial \zeta$ is the $(p + 1) \times T$ first derivative matrix of μ_i with regard to ζ. These first-order derivatives are available from Exercise 7.5.

Note that the GQL estimating equation (7.151) is a proper unbiased estimating equation for the zero vector and hence its solution, $\hat{\zeta}_{GQL}$, say, will be consistent. Furthermore, as the covariance matrix Σ_i is used for the weight matrix, to construct the GQL estimating equation (7.151), $\hat{\zeta}_{GQL}$ will also be more efficient than the moment estimator of ζ, for example. However, because θ is the dynamic dependence of $y_{i,t-1}$ on y_{it}, GQL estimation of this parameter by solving (7.151) may still produce some biases, especially in the finite sample case. A simulation study conducted by Sutradhar and Farrell (2007) supports this argument. For convenience, we present here a part of the simulation results from their study.

Performance of GQL Estimates Through Simulations

Consider the following simulation design from Sutradhar and Farrell (2007, Design 2, p. 458) with $K = 100$, $T = 4$, and $p = 2$.

$x_{it1} = 1.0$ $(t = 1, 2)$ and $x_{it1} = 0.0$ $(t = 3, 4)$ for $i = 1, \ldots, 25$,
$x_{it1} = 1.0$ for $i = 26, \ldots, 75$ and $t = 1, \ldots, 4$,
$x_{it1} = 0.0$ $(t = 1, 2)$ and $x_{it1} = 1.0$ $(t = 3, 4)$ for $i = 76, \ldots, 100$,
$x_{it2} = t/4$ for $i = 1, \ldots, 100$ and $t = 1, \ldots, 4$.

For $\beta_1 = \beta_2 = 1.0$ and $\theta = -3.0, -1.0, 0.0,$ and 1.0, we generate the data for 5000 times by using the model (7.142) and estimate the parameters β and θ by solving the GQL estimating equation (7.151). The simulated estimates are given in Table 7.8. The table also contains the estimated standard errors of the GQL estimates computed by using the asymptotic covariance expression

$$\text{cov}(\hat{\zeta}_{GQL}) = \left[\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \zeta} \Sigma_i^{-1} \frac{\partial \mu_i}{\partial \zeta'}\right]^{-1}. \quad (7.152)$$

Table 7.8 Simulated means, simulated standard errors, and estimated standard errors (in brackets following the SSEs), for the estimators of model (7.142) parameters under GQL, with $\beta_1 = \beta_2 = 1$, based on 5000 simulations.

θ	Method	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}$
−3.0	GQL	SM	1.0115	1.0169	−3.0334
		SSE (ESE)	0.2603 (0.2548)	0.4113 (0.4041)	0.5595 (0.5712)
−1.0	GQL	SM	1.0167	1.0100	−1.0048
		SSE (ESE)	0.2228 (0.2210)	0.5004 (0.4897)	0.5259 (0.5192)
0.0	GQL	SM	1.0236	0.9804	0.0510
		SSE (ESE)	0.2271 (0.2279)	0.6826 (0.6458)	0.6981 (0.6469)
1.0	GQL	SM	1.0387	0.9853	1.1091
		SSE (ESE)	0.2773 (0.2723)	0.9284 (0.8838)	0.9818 (0.9537)

The results of Table 7.8 show that in general the GQL estimates of β_1 and β_2 are almost unbiased, with an indication that for nonnegative values of $\theta = 0.0, 1.0$, the estimates are slightly biased. For these parameter values, the GQL estimate of the dynamic dependence parameter θ appear to be significantly biased. For example, the GQL estimate for $\theta = 1.0$ is shown to be 1.1091 which is highly biased.

In the next section, we consider an optimal GQL (OGQL) method where the GQL estimating equation is constructed by using both first-order and second-order responses. As indicated earlier, the second-order product responses must be more informative for the θ parameter as it defines the dynamic dependence of $y_{i,t-1}$ on y_{it}.

7.7.2.2.2 OGQL Estimation

Note that in GQL estimation by (7.151), the dynamic dependence parameter θ in (7.142) has been considered as a regression parameter similar to β. However, because $y_{i,t-1}$ is a different regression variable from the fixed effect covariate vector x_{it}, considering $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ as a basic statistic to construct the GQL estimating equation, does not appear to exploit sufficient information for the estimation of the model parameters, especially for θ. Thus to improve the GQL estimates of these parameters we construct a new GQL estimating equation by using all possible pairwise products and the first-order responses, instead of using only the first-order responses. For

$$y_i = (y_{i1}, \ldots, y_{iT})', \text{ and } s_i = (y_{i1}y_{i2}, \ldots, y_{iu}y_{it}, \ldots, y_{i,T-1}y_{iT})',$$

let

$$g_i = (y_i', s_i')', \tag{7.153}$$

denote the $T(T+1)/2 \times 1$ vector containing all first and distinct second-order paired responses. Suppose that

$$\nu_i = E[G_i] = (\mu_i', \lambda_i')', \tag{7.154}$$

where $\mu_i = E[Y_i]$ is the same as in the GQL estimating equation (7.151), and

$$\lambda_i = E(S_i) = E(Y_{i1}Y_{i2}, \ldots, Y_{iu}Y_{it}, \ldots, Y_{i,T-1}Y_{iT})'.$$

Note that $E(Y_{iu}Y_{it})$ is easily computed from (7.149) as

$$E(Y_{iu}Y_{it}) = \lambda_{iut} = \sigma_{iut} + \mu_{iu}\mu_{it}.$$

Furthermore, let $\Omega_i = \text{cov}(G_i)$ such that

$$\Omega_i = \begin{bmatrix} \text{cov}(Y_i) & \text{cov}(Y_i, S_i') \\ & \text{cov}(S_i) \end{bmatrix} = \begin{bmatrix} \Sigma_i & \Delta_i \\ & \Phi_i \end{bmatrix}, \tag{7.155}$$

is the $T(T+1)/2$ x $T(T+1)/2$ covariance matrix of the $T(T+1)/2$-dimensional extended vector f_i. By following (7.151), we now write an improved GQL estimating equation for $\zeta = (\beta', \theta)'$ given by

$$\sum_{i=1}^{K} \frac{\partial \nu_i'}{\partial \zeta} \Omega_i^{-1} (g_i - \nu_i) = 0, \tag{7.156}$$

for the estimation of both β and θ parameters. Note that as argued in Exercise 7.6, g_i in the estimating equation (7.156) is in fact a vector of sufficient statistics under the lag 1 dynamic model (7.142). For this reason, we refer to (7.156) as an OGQL estimating equation for $\zeta = (\beta', \theta)'$.

Note that under a quadratic exponential model for correlated binary data, Zhao and Prentice (1990), for example, have used an estimating equation similar to (7.156) for the estimation of the mean and covariance vector. Their model [see also Prentice (1988)] ignores the higher-order moments (more than second-order) and hence in their approach one cannot compute the fourth-order moment matrix Ω_i needed to construct the estimating equation (7.156). Moreover, under the present model, the covariances of the data (7.149) are functions of β and θ only. Thus, as opposed to Zhao and Prentice (1990) in the present setup one needs to compute fewer parameters.

As far as the computation of the Ω_i matrix in (7.156) is concerned, Zhao and Prentice (1990) have done this by using a 'working' normality based approach. To be specific, they compute this fourth-order moment matrix by pretending that the data follow a normal distribution with correct mean and variances computed under the binary quadratic model, even though in reality the data are binary. This 'working' assumption may not improve the efficiency [Sutradhar (2003)] of the estimates as compared to the 'working' independence assumption for the correlated binary data, which makes their approach less useful in practice where one needs to compute consistent as well as efficient estimates.

Computation of the Weight Matrix Ω_i

For convenience, we compute this matrix in three parts as follows.

Computation of $\Sigma_i = E[Y_i]$

The diagonal elements (σ_{itt}) of this matrix $\Sigma_i = (\sigma_{iut})$ are computed by (7.148) and the formulas for off-diagonal elements (σ_{iut}) are given in (7.149).

Computation of $\Delta_i = \mathbf{cov}[Y_i, S_i']$

Note that Δ_i is a third-order moment matrix. The elements of this matrix may be computed by using the formula for

$$\tilde{\delta}_{iuvt} = \text{cov}[Y_{iu}, Y_{iv}Y_{it}] = E[Y_{iu}Y_{iv}Y_{it}] - \mu_{iu}\lambda_{ivt}, \tag{7.157}$$

where μ_{it} is given by (7.145) and λ_{ivt} has the formula as in (7.154) [see also (7.149)]. Further note that for either $u = v$ or $u = t$, the third-order expectation $E[Y_{iu}Y_{iv}Y_{it}]$ in (7.157) reduces to the second-order expectation such as

$$E[Y_{iu}Y_{iv}Y_{it}] = E[Y_{iu}^2 Y_{it}] = E[Y_{iu}Y_{it}] = \lambda_{iut}, \text{ for } v = u,$$

which is known by (7.154). Thus, to complete the computation for all δ_{iuvt} elements we need to compute the third-order expectation $E[Y_{iu}Y_{iv}Y_{it}]$ only for distinct u, v, and t. The formula for this expectation is given by

$$\begin{aligned} E[Y_{iu}Y_{iv}Y_{it}] &= Pr(y_{iu} = 1, y_{iv} = 1, y_{it} = 1) \\ &= \sum_{S_1^*} \left[f(y_{i1}) \prod_{t=2}^{T} f(y_{it}|y_{i,t-1}) \right]_{y_{iu}=1, y_{iv}=1, y_{it}=1} \\ &= \delta_{iuvt}, \text{ (say)}, \end{aligned} \tag{7.158}$$

where $\sum_{S_1^*}$ indicates the summation over all $y_{ij} = 0, 1$ for $j \neq u, v, t$. In (7.158),

$$f(y_{i1}) = \mu_{i1}{}^{y_{i1}}[1 - \mu_{i1}]^{1-y_{i1}} \text{ and } f(y_{it}|y_{i,t-1}) = (p_{it|t-1})^{y_{it}}(1 - p_{it|t-1})^{1-y_{it}}.$$

Computation of $\Phi_i = \mathbf{cov}[S_i, S_i']$

To compute the elements of this fourth-order moments matrix, we write

$$\tilde{\phi}_{iuv\ell t} = E[Y_{iu}Y_{iv}Y_{i\ell}Y_{it}] - \lambda_{iuv}\lambda_{i\ell t}. \tag{7.159}$$

Note that for either $u = \ell$ or $v = \ell$, for example, the fourth-order expectation $E[Y_{iu}Y_{iv}Y_{i\ell}Y_{it}]$ in (7.159) reduces to the third-order expectation such as

$$E[Y_{iu}Y_{iv}Y_{i\ell}Y_{it}] = E[Y_{iu}^2 Y_{iv}Y_{it}] = E[Y_{iu}Y_{iv}Y_{it}] = \delta_{iuvt}, \text{ for } \ell = u,$$

which is known by (7.158). Similarly, when $u = \ell$ and $v = t$, for example, the fourth-order expectation $E[Y_{iu}Y_{iv}Y_{i\ell}Y_{it}]$ in (7.159) reduces to the second-order expectation such as

$$E[Y_{iu}Y_{iv}Y_{i\ell}Y_{it}] = E[Y_{iu}^2 Y_{iv}^2] = E[Y_{iu}Y_{iv}] = \lambda_{iuv}, \text{ for } \ell = u, \text{ and } t = v,$$

which is known by (7.154). Thus, to complete the computation for all $\phi_{iuv\ell t}$ elements we need to compute the fourth-order expectation $E[Y_{iu}Y_{iv}Y_{i\ell}Y_{it}]$ only for distinct u, v, ℓ, and t. The formula for this expectation is given by

$$\begin{aligned} E[Y_{iu}Y_{iv}Y_{i\ell}Y_{it}] &= Pr(y_{iu} = 1, y_{iv} = 1, y_{i\ell} = 1, y_{it} = 1) \\ &= \sum_{S_2^*} \left[f(y_{i1}) \prod_{t=2}^{T} f(y_{it}|y_{i,t-1}) \right]_{y_{iu}=1, y_{iv}=1, y_{i\ell}=1, y_{it}=1}, \end{aligned} \quad (7.160)$$

where $\sum_{S_2^*}$ indicates the summation over all $y_{ij} = 0,1$ for $j \neq u, v, \ell, t$.

Computation of the Derivatives $\partial v_i'/\partial \zeta$

To construct the OGQL estimating equation (7.156), we also need to compute the first-order derivatives of $v_i' = (\mu_i', \lambda_i')$ with respect to $\zeta = (\beta', \theta)'$. The derivatives of μ_i with respect to β and θ are available from Exercise 7.5. Now to compute $\partial \lambda_i'/\partial \zeta$, it is convenient to write that for $u < t$

$$\lambda_{iut} = E(Y_{iu}Y_{it}) = \sum_{y_{iu}, y_{it} \notin S^*} \left[f(y_{i1}) \prod_{t=2}^{T} f(y_{it}|y_{i,t-1}) \right]_{y_{iu}=1, y_{it}=1}, \quad (7.161)$$

where $\sum_{y_{iu}, y_{it}} \notin S^*$ reflects the summation over all components of y_i except y_{iu} and y_{it}. Then, using (7.161), it is sufficient to determine

$$\begin{aligned} \partial \lambda_{iut}/\partial \beta = \sum_{y_{iu}, y_{it} \notin S^*} & [f(y_{i1}) \sum_{k=2}^{T} \{\prod_{j \neq k}^{T} f(y_{ij}|y_{i,j-1})\} f(y_{ik}|y_{i,k-1})(y_{ik} - p_{ik|k-1})x_{ik} \\ & + \{\textstyle\prod_{j=2}^{T} f(y_{ij}|y_{i,j-1})\} f(y_{i1})(y_{i1} - \mu_{i1})x_{i1}]_{y_{iu}=1, y_{it}=1}, \end{aligned}$$

or equivalently,

$$\begin{aligned} \frac{\partial \lambda_{iut}}{\partial \beta} &= \sum_{y_{iu}, y_{it} \notin S^*} [f(y_{i1}) \prod_{j=2}^{T} f(y_{ij}|y_{i,j-1}) \\ &\quad \times \{\sum_{k=2}^{T} (y_{ik} - p_{ik|k-1})x_{ik} + (y_{i1} - \mu_{i1})x_{i1}\}]_{y_{iu}=1, y_{it}=1}, \end{aligned} \quad (7.162)$$

where $p_{ik|k-1}$ is given in (7.142). Similarly,

$$\frac{\partial \lambda_{iut}}{\partial \theta} = \sum_{y_{iu},y_{it} \notin S^*} [f(y_{i1}) \prod_{j=2}^{T} f(y_{ij}|y_{i,j-1}) \{ \sum_{k=2}^{T} (y_{ik} - p_{ik|k-1}) y_{i,k-1} \}]_{y_{iu}=1,y_{it}=1}. \tag{7.163}$$

Performance of the OGQL Estimates Through a Simulation Study

By using the same parameters and time-dependent covariates as in the simulation study in the last section, we now have obtained simulated OGQL estimates for β and θ, by solving the OGQL estimating equation in (7.156). The simulated mean and standard errors of the estimates are shown in Table 7.9. The estimated standard errors computed by using the asymptotic covariance formula

$$\text{cov}(\hat{\zeta}_{OGQL}) = \left[\sum_{i=1}^{K} \frac{\partial \nu_i'}{\partial \zeta} \Omega_i^{-1} \frac{\partial \nu_i}{\partial \zeta'} \right]^{-1}, \tag{7.164}$$

are also given in the same table. The results of the table indicate a substantial improvement over the GQL estimates shown in Table 7.8. For example, when $\theta = 1.0$, the OGQL estimates for $\beta_1 = 1.0$, $\beta_2 = 1.0$, and θ are

$$1.0085,\ 1.0591,\ \text{and}\ 0.9832,$$

respectively, whereas the corresponding GQL estimates from Table 7.8 are

$$1.0387,\ 0.9853,\ \text{and}\ 1.1091.$$

Thus the OGQL estimates are much less biased than the GQL estimates, showing a large improvement, especially for the estimation of the dynamic dependence parameter θ. Also, the standard errors of the estimates are much smaller under the OGQL approach as compared to the GQL approach.

Table 7.9 Simulated means, simulated standard errors, and estimated standard errors (in brackets following the SSEs), under OGQL (and ML) method with $\beta_1 = \beta_2 = 1$, based on 5000 simulations.

θ	Method	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}$
−3.0	OGQL	SM	1.0142	1.0250	−3.0566
		SSE (ESE)	0.2425 (0.2425)	0.3346 (0.3306)	0.2952 (0.2933)
−1.0	OGQL	SM	1.0119	1.0308	−1.0327
		SSE (ESE)	0.2154 (0.2138)	0.3411 (0.3371)	0.2642 (0.2670)
0.0	OGQL	SM	1.0105	1.0351	−0.0203
		SSE (ESE)	0.2229 (0.2249)	0.4121 (0.4011)	0.3185 (0.3113)
1.0	OGQL	SM	1.0085	1.0591	0.9832
		SSE (ESE)	0.2569 (0.2595)	0.5246 (0.5153)	0.4042 (0.4017)

7.7.2.2.3 Likelihood Estimation

Note that unlike in the longitudinal setup for count data, the likelihood estimation under the present binary dynamic logit model is quite manageable, in fact it is much easier than the OGQL estimation. However, when this model is extended to accommodate random effects, the maximum likelihood estimation will be more complicated than the OGQL (or GQL) approach for the estimation of the parameters of such a dynamic mixed model. These issues are discussed in detail in the next two chapters.

Turning back to the likelihood estimation for the present BDL model (7.142), by using the conventional notation $y_{i0} = 0$, the likelihood may be written as

$$\begin{aligned} L(\beta,\theta) &= \prod_{i=1}^{K}\left[\frac{\exp[(x_{i1}'\beta)y_{i1}]}{1+\exp(x_{i1}'\beta)}\prod_{t=2}^{T}\frac{\exp[(x_{it}'\beta+\theta y_{i,t-1})y_{it}]}{1+\exp(x_{it}'\beta+\theta y_{i,t-1})}\right] \\ &= \prod_{i=1}^{K}[g_{i1}]\prod_{t=2}^{T}\left[g_{it|t-1}\right], \text{ (say)} \end{aligned} \tag{7.165}$$

(see also Exercise 7.6). One may then write the log-likelihood function as

$$\log L = \sum_{i=1}^{I}\sum_{t=1}^{T} y_{it}(x_{it}'\beta+\theta y_{i,t-1}) - \sum_{i=1}^{I}\sum_{t=1}^{T}\log[1+\exp(x_{it}'\beta+\theta y_{i,t-1})], \tag{7.166}$$

that yields the likelihood estimating equations for β and θ given by

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{I}\sum_{t=1}^{T}[y_{it}-p_{it|t-1}]x_{it}' = 0, \tag{7.167}$$

and

$$\frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^{I}\sum_{t=1}^{T}[y_{it}-p_{it|t-1}]y_{i,t-1} = 0, \tag{7.168}$$

where $p_{it|t-1} = \exp(x_{it}'\beta+\theta y_{i,t-1})/[1+\exp(x_{it}'\beta+\theta y_{i,t-1})]$.

Performance of the ML Estimates Through a Simulation Study

It appears from the likelihood function (7.165) that the first-order responses $\{y_{it}\}$ and the second-order responses $\{y_{it}y_{i,t-1}\}$ must be sufficient for the estimation of the parameters β and θ. Because the OGQL approach also uses these first$-$ and second-order responses, the ML and OGQL approaches may yield the same estimates for the present BDL model. In fact, the simulation study conducted by Sutradhar and Farrell (2007) supports this observation, where it is reported that the ML and OGQL estimates are the same. For this reason, it is not necessary to produce any separate table with results on ML estimation, rather we have indicated in Table 7.9 that ML estimates are the same as the OGQL estimates.

Note that the estimated standard errors of the ML estimates obtained from (7.167) – (7.168), were computed by using covariance matrix

$$\mathrm{cov}[\hat{\zeta}_{ML}] = Q^{-1}, \tag{7.169}$$

where the so-called Fisher information matrix Q, is computed as

$$Q = \begin{bmatrix} -E(\frac{\partial^2 \log L}{\partial \beta \beta'}) & -E(\frac{\partial^2 \log L}{\partial \beta \partial \theta}) \\ & -E(\frac{\partial^2 \log L}{\partial \theta^2}) \end{bmatrix}. \tag{7.170}$$

7.7.2.3 Fitting Asthma Data to the BDL Model: An Illustration

As an illustration of the application of GQL and OGQL (=ML) approaches, we consider a dataset that contains complete records of I = 537 children from Steubenville, Ohio, each of whom was examined annually at ages 7 through 10. The repeated response is the asthma status (1 = yes, 0 = no) of a child on each of the T = 4 occasions. Maternal smoking status was considered as a covariate; it was recorded as 1 if the mother smoked regularly, and 0 otherwise. The dataset is given in Table 7F in the appendix. It is of interest to estimate the dynamic dependence parameter that explains how the asthma status at a given time is affected by the previous asthma status. It is also of interest to compute the effect of smoking by the mother on the asthma status of her child.

Note that this dataset was earlier analyzed by Zeger, Liang and Albert (1988), Sutradhar (2003), and Sutradhar and Farrell (2007), among others. As the binary responses for each child are repeatedly collected over a period of T = 4 years, it is likely that they will be longitudinally correlated. Sutradhar (2003) has modelled the longitudinal correlations by using a $T \times T$ stationary autocorrelation structure (7.65), and obtained the regression estimates $\hat{\beta}_1$ (intercept) = -1.820 and $\hat{\beta}_2$ (maternal smoking effect) = 0.263, by solving the GQL estimating equation (7.66). The stationary lag correlations were estimated by (7.67), and they were found to be $\hat{\rho}_1$ = 0.397, $\hat{\rho}_2$ = 0.310, and $\hat{\rho}_3$ = 0.297, respectively. Unlike Sutradhar (2003), Sutradhar and Farrell (2007) fitted the BDL model (7.142) to the same asthma data . Thus, Sutradhar and Farrell (2007) estimated θ, the lag 1 dependence parameter, whereas Sutradhar (2003) computed three lag correlations, the lag 1 correlation being similar to but different from θ. But, β_1 and β_2 denote the regression effects of the same two covariates both in Sutradhar (2003) and Sutradhar and Farrell (2007). Note, however, that these regression effects influence the means of the response variable under the BDL model (7.142) in a different way from that of the model considered by Sutradhar (2003), means are being nonstationary and dynamic under the model (7.142). To be more specific, the expected asthma status of a child at a given year is influenced by the history of the covariates under the BDL model such as the history of the smoking habits of the parents in a household, whereas in the existing literature such as in Sutradhar (2003), the expected asthma status at a given year is influenced only by the smoking habit of the parents during that specified year.

In Table (7.10), we reproduce the GQL (7.151) and OGQL (≅ML) (7.156) estimates of the parameters of the BDL model from Sutradhar and Farrell (2007). The standard errors of the estimates are also given.

Table 7.10 For the wheeze data where $T = 4$, estimates obtained using GQL (7.151) and OQGL (=ML) (7.156) for the binary dynamic logit model (7.142) containing a lag 1 dependence parameter. For each estimation approach, the estimated covariance matrix is given below the estimates for β_1, β_2, and θ.

Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$
GQL	−1.7738	0.2842	−0.4943
	1.40×10^{-2}	-3.48×10^{-3}	-1.08×10^{-1}
		1.60×10^{-2}	-3.33×10^{-2}
			1.45×10^{0}
OGQL (≅ML)	−2.1886	0.2205	1.9544
	7.94×10^{-3}	-6.69×10^{-3}	-5.25×10^{-3}
		1.75×10^{-2}	-3.66×10^{-4}
			2.35×10^{-2}

The results in Table 7.10 show that the GQL estimates appear to be different from the OGQL or ML estimates. In addition, the standard errors (computed from the diagonal elements of the estimated covariance matrix) of the GQL estimates are relatively larger than counterparts obtained under the OGQL or ML approach. This illustrates that, as expected, the OGQL or ML approach is more efficient than the GQL approach.

The GQL estimates for β_1 and β_2 are close to the corresponding GQL estimates found by Sutradhar (2003). However, the GQL estimates under the BDL model appear to be more efficient than those under the traditional longitudinal model considered by Sutradhar (2003), specifically, the standard error of $\hat{\beta}_2$ of 0.177 obtained by Sutradhar (2003), whereas the analogous standard error in Table 7.10 arrived at using GQL under the BDL model is 0.126. The standard errors of $\hat{\beta}_1$ are the same under both models. Under the BDL model, the OGQL estimates for β_1 and β_2 are more efficient than the GQL estimates. The estimates obtained for the regression parameters under OGQL are $\hat{\beta}_{1,OGQL} = -2.19$ and $\hat{\beta}_{2,OGQL} = 0.22$, which are generally different than their GQL counterparts: $\hat{\beta}_{1,GQL} = -1.77$ and $\hat{\beta}_{2,GQL} = 0.28$.

As far as the dynamic dependence parameter is concerned, the GQL approach produces a negative estimate, namely, $\hat{\theta}_{GQL} = -0.49$, whereas the OGQL or ML approach produces a high positive estimate, $\hat{\theta}_{OGQL} = 1.95$. The simulation study in the last section showed that the OGQL approach produces a reliable estimate for the dynamic dependence parameter θ, whereas the GQL estimate can be different from the true value. This leads one to accept the high positive estimate 1.95 for the θ parameter. Note that this high positive estimate of θ is in agreement with the positive lag 1 correlation estimate 0.397 found in Sutradhar (2003).

We now use the OGQL or ML estimates to interpret the data. The high positive value $\hat{\theta}_{OGQL} = 1.95$ shows that a previous asthma attack contributes highly to the asthma attack at a given time. The negative estimate for β_1, that is, $\hat{\beta}_{1,OGQL} = -2.19$ and the positive value for $\hat{\beta}_{2,OGQL} = 0.22$ indicate that even if there is an overall decreasing tendency in asthma attack rate, this rate, however, increases for the children whose mothers are smokers.

7.7.3 Application of the Binary Dynamic Logit (BDL) Model in an Adaptive Clinical Trial Setup

The BDL model considered in the last section is developed based on certain fixed covariates. In some practical situations such as in longitudinal clinical studies, it may happen that some of the covariates such as treatments are selected randomly following an adaptive design, whereas the rest of the covariates may be fixed by nature. For details on the construction of longitudinal design weights, we refer to Section 7.6.1.1. The purpose of this section is to discuss the effects of the design weights selection on the parameter estimation including the treatment effects, after taking the longitudinal correlations of the repeated binary responses into account. Note that with regard to the longitudinal correlation structure, it was assumed in Section 7.6 that once the treatment was selected for the ith $(i = 1,\ldots,K)$ patient, the repeated binary responses follow a stationary AR(1) correlation structure (see (7.100)). However, here we assume that the repeated binary responses follow the binary dynamic logit model (7.142). Thus, the correlations modelled through the dynamic dependence parameter θ can be nonstationary.

7.7.3.1 Random Treatments Based BDL Model

This model was developed by Sutradhar and Jowaheer (2009). Let y_{it} denote the binary response for the ith $(i = 1,\ldots,K)$ individual collected at time t $(t = 1,\ldots,T)$, $x_{it} = [x_{it1},\ldots,x_{itu},\ldots,x_{itp}]'$ be the p-dimensional vector of time-dependent fixed covariates, and $\delta_i = [\delta_{i1},\ldots,\delta_{ij},\ldots,\delta_{ic}]'$ be the c-dimensional random indicator vector that determines the selection of one treatment for the ith individual out of $c+1$ treatments. In Section 7.6, we considered $c+1 = 2$, for simplicity. Furthermore, let $\alpha = [\alpha_1,\ldots,\alpha_j,\ldots,\alpha_c]'$ and $\beta = [\beta_1,\ldots,\beta_u,\ldots,\beta_p]'$ denote the effects of δ_i and x_{it}, respectively, on the binary response y_{it}. This is of interest to estimate α and β consistently and efficiently.

For $j = 1,\ldots,c$, suppose that

$$Pr[\delta_{i1} = 0,\ldots,\delta_{i,j-1} = 0, \delta_{ij} = 1, \delta_{i,j+1} = 0,\ldots,\delta_{i,c} = 0] = w_{ij} \qquad (7.171)$$

is the probability of the selection of the jth treatment for the ith patient. It then follows that the probability for the selection of $(c+1)$th treatment for the assignment of the ith patient is given by

$$Pr[\delta_{i1}=0,\ldots,\delta_{i,j-1}=0,\delta_{ij}=0,\delta_{i,j+1}=0,\ldots,\delta_{i,c}=0]=1-\sum_{j=1}^{c}w_{ij}. \quad (7.172)$$

In adaptive clinical trials, these probabilities w_{ij}s are referred to as the adaptive design weights and they are computed based on a suitable scheme such as the simple longitudinal play-the-winner rule, discussed in Section 7.6.1.1, using two treatments.

Note that as far as the correlation structure for the repeated binary responses $y_{i1},\ldots,y_{it},\ldots,y_{iT}$, is concerned, we follow the BDL model (7.142), and re-express it conditional on the random treatments as follows.

$$p^*_{i1} = Pr[y_{i1}=1|\delta_i] = \frac{\exp(\delta_i'\alpha + x_{i1}'\beta)}{1+\exp(\delta_i'\alpha + x_{i1}'\beta)}$$

$$p^*_{it|t-1} = P(y_{it}=1|y_{i,t-1},\delta_i) = \frac{\exp(\delta_i'\alpha + x_{it}'\beta + \theta y_{i,t-1})}{1+\exp(\delta_i'\alpha + x_{it}'\beta + \theta y_{i,t-1})} \quad (7.173)$$

[see also Amemiya (1985); Zhao and Prentice (1990); Aitkin and Alfo (1998)] for $t=2,\ldots,T$, where θ is the dynamic dependence parameter, α is the effect of the treatment indicator vector $\delta_i=[\delta_{i1},\ldots,\delta_{ij},\ldots,\delta_{ic}]'$, and β is the effect of the time-dependent prognostic fixed covariates x_{it}.

Note that in practice δ_i vectors for $i=1,\ldots,K$, are unknown. In adaptive clinical trial studies, they are usually generated by using the design weights w_{ij} which are, however, known based on a randomized scheme such as the SLPW rule. More specifically, δ_i can be generated by following the multinomial distribution given by

$$Pr[\delta_{i1},\ldots,\delta_{i,c}] = \frac{1!}{\delta_{i1}!\ldots\delta_{i,c}!(1-\sum_{j=1}^{c}\delta_{ij})!} w_{i1}^{\delta_{i1}}\cdots w_{i,c}^{\delta_{i,c}}(1-\sum_{j=1}^{c}w_{ij})^{1-\sum_{j=1}^{c}\delta_{ij}}. \quad (7.174)$$

In (7.174), all of the c variables δ_{ij} for $j=1,\ldots,c$ may be 0 or at most one of them may assume the value 1 so that $\sum_{j=1}^{c}\delta_{ij}=0$ or 1. When all $\delta_{ij}=0$, for $j=1,\ldots,c$, the physician selects the $c+1$th treatment for the individual.

7.7.3.1.1 Unconditional Moments Up to Order Four

Let

$$\pi_{it}(\alpha_j) = \frac{\exp(\alpha_j + x_{it}'\beta)}{1+\exp(\alpha_j + x_{it}'\beta)} \text{ and } \tilde{p}_{it}(\alpha_j) = \frac{\exp(\alpha_j + x_{it}'\beta + \theta)}{1+\exp(\alpha_j + x_{it}'\beta + \theta)}.$$

It then follows that

$$\mu_{it} = E[Y_{it}] = Pr[Y_{it} = 1] = \sum_{j=1}^{c} w_{ij}\mu_{it}(\alpha_j) + (1 - \sum_{j=1}^{c} w_{ij})\mu_{it}(0), \tag{7.175}$$

with $\mu_{it}(0) = [\mu_{it}(\alpha_j)]_{|\alpha_j=0}$, and where by similar calculations as in (7.145) one writes

$$\mu_{it}(\alpha_j) = \pi_{it}(\alpha_j) + \mu_{i,t-1}(\alpha_j)[\tilde{p}_{it}(\alpha_j) - \pi_{it}(\alpha_j)], \text{ for } t = 2,\ldots,T, \tag{7.176}$$

with $\mu_{i1}(\alpha_j) = \pi_{i1}(\alpha_j)$.

To compute the second− and higher-order moments up to order four, conditional on δ_i, we first define

$$g^*_{i1}(\delta_i) = \frac{\exp[(\delta_i'\alpha + x_{i1}'\beta)y_{i1}]}{1+\exp(\delta_i'\alpha + x_{i1}'\beta)},$$

$$g^*_{it|t-1}(\delta_i) = \frac{\exp[(\delta_i'\alpha + x_{it}'\beta + \theta y_{i,t-1})y_{it}]}{1+\exp(\delta_i'\alpha + x_{it}'\beta + \theta y_{i,t-1})}, \text{ for } t = 2,\ldots,T. \tag{7.177}$$

Note that when δ_i has 1 in the jth position, one may write these functions as

$$g^*_{i1}(\alpha_j) = \frac{\exp[(\alpha_j + x_{i1}'\beta)y_{i1}]}{1+\exp(\alpha_j + x_{i1}'\beta)}, \quad g^*_{it|t-1}(\alpha_j) = \frac{\exp[(\alpha_j + x_{it}'\beta + \theta y_{i,t-1})y_{it}]}{1+\exp(\alpha_j + x_{it}'\beta + \theta y_{i,t-1})},$$

for $t = 2,\ldots,T$.

For $u < v < l < t$, one may then compute the second, third, and fourth order unconditional moments by using the corresponding conditional moments given by

$$\begin{aligned}\lambda^*_{ut}(\delta_i) &= E[Y_{iu}Y_{it}|\delta_i] \\ &= \Sigma_{S_1^*}[(g^*_{i1}(\delta_i)\prod_{t=2}^{T} g^*_{i,t|t-1})(\delta_i)_{y_{iu}=1,y_{it}=1}],\end{aligned} \tag{7.178}$$

$$\begin{aligned}\delta^*_{uvt}(\delta_i) &= E[Y_{iu}Y_{iv}Y_{it}|\delta_i] \\ &= \Sigma_{S_2^*}[(g^*_{i1}(\delta_i)\prod_{t=2}^{T} g^*_{i,t|t-1})(\delta_i)_{y_{iu}=1,y_{iv}=1,y_{it}=1}],\end{aligned} \tag{7.179}$$

and

$$\begin{aligned}\phi^*_{uvlt}(\delta_i) &= E[Y_{iu}Y_{iv}Y_{il}Y_{it}|\delta_i] \\ &= \Sigma_{S_3^*}[(g^*_{i1}(\delta_i)\prod_{t=2}^{T} g^*_{i,t|t-1})(\delta_i)_{y_{iu}=1,y_{iv}=1,y_{il}=1,y_{it}=1}],\end{aligned} \tag{7.180}$$

respectively, where $\Sigma_{S_1^*}$ indicates the summation over all $y_{ik} = 0,1$ for $k \neq u,t$, and similarly $\Sigma_{S_2^*}$ and $\Sigma_{S_3^*}$ reflect the summation over all $y_{ik} = 0,1$ for $k \neq u,v,t$, and over all $y_{ik} = 0,1$ for $k \neq u,v,l,t$, respectively. The unconditional second−, third−, and fourth-order moments have the formulas as

$$\begin{aligned} E(Y_{iu}Y_{it}) &= \lambda_{iut} \\ &= E_{\delta_i}E[(Y_{iu}Y_{it})|\delta_i] \\ &= \sum_{j=1}^{c} w_{ij}[\Sigma_{S_1^*}(g_{i1}^*(\alpha_j)\prod_{t=2}^{T} g_{i,t|t-1}^*(\alpha_j))_{y_{iu}=1,y_{it}=1}] \\ &\quad +(1-\sum_{j=1}^{c} w_{ij})[\Sigma_{S_1^*}(g_{i1}^*(0)\prod_{t=2}^{T} g_{i,t|t-1}^*(0))_{y_{iu}=1,y_{it}=1}], \end{aligned} \quad (7.181)$$

$$\begin{aligned} E(Y_{iu}Y_{iv}Y_{it}) &= \delta_{iuvt}^* \\ &= E_{\delta_i}E[(Y_{iu}Y_{iv}Y_{it})|\delta_i] \\ &= \sum_{j=1}^{c} w_{ij}[\Sigma_{S_2^*}(g_{i1}^*(\alpha_j)\prod_{t=2}^{T} g_{i,t|t-1}^*(\alpha_j))_{y_{iu}=1,y_{iv}=1,y_{it}=1}] \\ &\quad +(1-\sum_{j=1}^{c} w_{ij})[\Sigma_{S_2^*}(g_{i1}^*(0)\prod_{t=2}^{T} g_{i,t|t-1}^*(0))_{y_{iu}=1,y_{iv}=1,y_{it}=1}], \end{aligned} \quad (7.182)$$

and

$$\begin{aligned} E(Y_{iu}Y_{iv}Y_{il}Y_{it}) &= \phi_{iuvlt}^* \\ &= E_{\delta_i}E[(Y_{iu}Y_{iv}Y_{il}Y_{it})|\delta_i] \\ &= \sum_{j=1}^{c} w_{ij}\left[\Sigma_{S_3^*}(g_{i1}^*(\alpha_j)\prod_{t=2}^{T} g_{i,t|t-1}^*(\alpha_j))_{y_{iu}=1,y_{iv}=1,y_{il}=1,y_{it}=1}\right] \\ &\quad +(1-\sum_{j=1}^{c} w_{ij})[\Sigma_{S_3^*}(g_{i1}^*(0) \\ &\quad \times\prod_{t=2}^{T} g_{i,t|t-1}^*(0))_{y_{iu}=1,y_{iv}=1,y_{il}=1,y_{it}=1}], \end{aligned} \quad (7.183)$$

respectively. Note that as T is small in the longitudinal setup, such as T = 3 or 4, the expectations in (7.181), (7.182), and (7.183) are easily evaluated.

7.7.3.1.2 Extended WGQL (EWGQL) or Weighted OGQL (WOGQL) Estimating Equation

Following the OGQL estimating equation (7.156), we now write the design weights ($w \equiv \{w_{ij}\}$) based OGQL estimating equation for

$$\zeta^* = (\alpha', \beta', \theta)'$$

as

$$\sum_{i=1}^{K} \frac{\partial \nu_i'(w)}{\partial \zeta^*} \Omega_i^{-1}(w)(g_i - \nu_i(w)) = 0, \tag{7.184}$$

where

$$g_i = (y_i', s_i')'$$

with

$$y_i = (y_{i1}, \ldots, y_{iT})', \text{ and } s_i = (y_{i1}y_{i2}, \ldots, y_{iu}y_{it}, \ldots, y_{i,T-1}y_{iT})'.$$

Note that the design weights based formula for

$$\nu_i(w) = E[G_i(w)]$$

can be computed by (7.175) and (7.181), and

$$\Omega_i(w) = \text{cov}[G_i] = \begin{bmatrix} \text{cov}(Y_i) & \text{cov}(Y_i, S_i') \\ & \text{cov}(S_i) \end{bmatrix} = \begin{bmatrix} \Sigma_i(w) & \Delta_i(w) \\ & \Phi_i(w) \end{bmatrix}, \tag{7.185}$$

can be computed by using $(7.181)-(7.183)$. Note that in view of the adaptive design weights based WGQL estimating equation (7.119) for β, the weighted OGQL (WOGQL) estimating equation in (7.184) may also be referred to as an extended WGQL (EWGQL) estimating equation for α, β, and θ parameters. The derivatives in (7.184) require some lengthy but straightforward algebra. They are available from the appendix in Sutradhar and Jowaheer (2009).

Performance of the EWGQL Approach: A Simulation Study

A part of the simulation results from Sutradhar and Jowaheer (2009) is given here to demonstrate the performance of the EWGQL estimating equation (7.184) for the dynamic dependence and regression parameters including the treatment effects. This is done for known adaptive design weights. Three different combinations of design weights are considered: (1) equal weights, (2) decreasing weights, the largest weight being assigned for the selection of the best treatment, and also (3) increasing weights, the smallest weight being assigned for the best treatment. The simulation design including the parameter values is chosen as follows.

$K = 200$ individuals, $c+1 = 3$ treatments, $T = 4$ time points;
Equal weights: $w_{i1} = 0.33, w_{i2} = 0.33$, and $w_{i3} = 0.34$, for $i = 1, \ldots, 200$;
Decreasing weights: $w_{i1} = 0.60, w_{i2} = 0.30$, and $w_{i3} = 0.10$, for $i = 1, \ldots, 200$;

Increasing weights: $w_{i1} = 0.10, w_{i2} = 0.20$, and $w_{i3} = 0.70$, for $i = 1, \ldots, 200$;
Relative treatment effects: $\alpha_1 = 1.0$, and $\alpha_2 = 0.5$;
Two prognostic covariate effects: $\beta_1 = \beta_2 = 1.0$;
Dynamic dependence parameter $\theta = 1.0$;
$x_{it3} = 1.0\ (t = 1, 2)$ and $x_{it3} = 0.0\ (t = 3, 4)$ for $i = 1, \ldots, 50$,
$x_{it3} = 1.0$ for $i = 51, \ldots, 150$ and $t = 1, \ldots, 4$,
$x_{it3} = 0.0\ (t = 1, 2)$ and $x_{it3} = 1.0\ (t = 3, 4)$ for $i = 151, \ldots, 200$;
$x_{it4} = t/4$ for $i = 1, \ldots, 200$ and $t = 1, \ldots, 4$.

Note that in the simulation study, we first use the design weights in the multinomial distribution (7.174) and generate 200 sets of values of the treatment covariates $(\delta_{i1},\ \delta_{i2})$. The treatment is not changed over time for an individual, thus these values are kept the same for all $t = 1, \ldots, 4$, and we use $x_{it1} = \delta_{i1}$, and $x_{it2} = \delta_{i2}$.

The repeated binary responses generated under a simulation, along with the covariate values, are used in (7.184) to obtain the EWGQL estimates for all five parameters, namely,

$$\alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2,\ \text{and}\ \theta.$$

The simulations are repeated 1000 times. For a given set of values for the design weights: $w_{i1},\ w_{i2},\ w_{i3}$, the average estimates, that is, the simulated means, along with their standard errors, for all five parameters, are computed. These results are given in Table 7.11.

Table 7.11 Simulated means, simulated standard errors, estimated standard errors, and mean squared errors for the estimators of model parameters under the EWGQL approach, with unequal treatment effects $\alpha_1 = 1.0,\ \alpha_2 = 0.5$; prognostic covariate effects $\beta_1 = \beta_2 = 1$; and large positive dynamic dependence parameter $\theta = 1.0$, based on 1000 simulations.

Treatment Effects $(\alpha_1,\ \alpha_2)$	Design Weights (w_{i1}, w_{i2}, w_{i3})	Quantity	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$
(1.0, 0.5)	(0.33, 0.33, 0.34)	SM	1.0285	0.5043	1.0002	1.0783	0.9706
		SSE	0.3142	0.2623	0.2241	0.5147	0.3540
		ESE	0.3116	0.2597	0.2252	0.5030	0.3536
		MSE	0.0995	0.0688	0.0502	0.2711	0.1262
	(0.60, 0.30, 0.10)	SM	1.0054	0.4921	1.0154	1.0889	0.9778
		SSE	0.3016	0.3218	0.2641	0.6173	0.3972
		ESE	0.3024	0.3128	0.2547	0.6115	0.3956
		MSE	0.0910	0.1036	0.0700	0.3889	0.1583
	(0.10, 0.20, 0.70)	SM	1.1362	0.5193	1.0032	1.0505	0.9844
		SSE	0.9354	0.3186	0.2008	0.4104	0.3117
		ESE	0.5020	0.3097	0.1965	0.3955	0.3036
		MSE	0.8935	0.1019	0.0403	0.1710	0.0974

The results of the table indicate that the EWGQL approach produces almost unbiased estimates for the parameters including the treatment effects, when equal weights $w_{i1} = 0.33,\ w_{i2} = 0.33,\ w_{i3} = 0.34$, are used to choose the treatments.

In this case, the simulated estimates for the treatment effects are 1.0285, 0.5043, with simulated standard errors 0.3142, 0.2623, respectively. The estimates for the other parameters including the dynamic dependence parameter also appear to be very close to their corresponding true values. It is further seen that the estimates of the treatment effects are better when larger design weights are considered for the selection of the first treatment. To be specific, when the first treatment is considered to be the best, the estimate of α_1 has the minimum MSE (0.0910), when $w_{i1} = 0.60$, $w_{i2} = 0.30$, and $w_{i3} = 0.10$. This result indicates that it is not only that there should be a provision for ethical reasons to assign the best treatment to most of the patients, in fact, this type of assignment also would help to estimate the treatment effect efficiently. By the same token, if the best treatment is assigned to a few patients, it can happen that the treatment effect may not be estimated unbiasedly, that is, consistently. For example, when the best treatment was assigned with probability weight $w_{i1} = 0.10$, followed by $w_{i2} = 0.20$ and $w_{i3} = 0.70$, the best treatment effect $\alpha = 1.0$ was estimated as $\hat{\alpha}_1 = 1.1362$ with a large bias.

Note that once an estimate is obtained for the true parameter value, in practice one may like to compute the standard error of the estimate mainly for the construction of a confidence interval at a desired level of significance. For the purpose, under each of the 1000 simulations, we have also computed the asymptotic standard errors of the estimates for the parameters α_1, α_2, β_1, β_2, and θ by using the formula

$$[\sum_{i=1}^{I} \frac{\partial \nu_i'}{\partial \zeta^*} \Omega_i^{-1} \frac{\partial \nu_i}{\partial \zeta^{*\prime}}]^{-1},$$

for the asymptotic covariance matrix obtained from (7.184). Next, the averages of these 1000 standard errors for each of the five estimates were computed, and reported as ESE in Table 7.11. It appears from the table that in general the ESE agrees with the SSE, provided the design weights are chosen reflecting the treatment effects. For example, for $\alpha_1 = 1.0$, $\alpha_2 = 0.5$, when weights are chosen as $w_{i1} = 0.6$, $w_{i2} = 0.3$, and $w_{i3} = 0.1$, in monotonic decreasing order, the ESEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\theta}$ are

$$0.3024,\ 0.3128,\ 0.2547,\ 0.6115,\ 0.3956,$$

respectively, which agree with the corresponding SSEs

$$0.3016,\ 0.3218,\ 0.2641,\ 0.6173, 0.3972.$$

However, when the design weights are not chosen reflecting the treatment effects, the ESE can be biased for SSE, and hence may not be reliable. For example, when $w_{i1} = 0.10$, $w_{i2} = 0.20$, and $w_{i3} = 0.70$, the results in the table show that the ESE of $\hat{\alpha}_1$ is 0.5020, whereas for this case SSE is 0.9354.

Exercises

7.1. (Section 7.2.1.2) [Nonstationary Bahadur bivariate binary distribution]
Use the $T = 3-$dimensional Bahadur's binary density (7.18) and show that y_{i1} and y_{i2} follow the bivariate density given by

$$f(y_{i1}, y_{i2}) = \Pi_{t=1}^{2} \pi_{it}^{y_{it}} (1-\pi_{it})^{1-y_{it}} \left[1 + c^*_{i,12} \frac{[(y_{i1}-\pi_{i1})(y_{i2}-\pi_{i2})]}{[\pi_{i1}(1-\pi_{i1})\pi_{i2}(1-\pi_{i2})]^{\frac{1}{2}}} \right]. \quad (7.186)$$

Also show that this bivariate density provides

$$\begin{aligned} E[Y_{it}] &= \pi_{it} \ \text{ for } t = 1,2 \\ E[Y_{i1}Y_{i2}] &= \pi_{i1}\pi_{i2} + c^*_{i,12}[\pi_{i1}(1-\pi_{i1})\pi_{i2}(1-\pi_{i2})]^{1/2}, \end{aligned} \quad (7.187)$$

yielding $c^*_{i,12}$ as the correlation between y_{i1} and y_{i2}.

7.2. (Section 7.2.2.2) [Higher lag autocovariances for nonstationary ODD model]
The lag 1 and 2 autocovariances are computed in $(7.31) - (7.32)$. For the computation of the higher (than 2) lag autocovariances, verify that for $w \geq 3$, the joint probabilities can be computed by using the formula

$$\begin{aligned} Pr(Y_{i,t-w} = 1,\, Y_{it} = 1) = {} & Pr(Y_{i,t-w} = 1) \\ & \times \sum_{j_1=0}^{1} \cdots \sum_{j_{w-1}=0}^{1} Pr(Y_{i,t-(w-1)} = j_1 | Y_{i,t-w} = 1) \\ & \times \left\{ \Pi_{r=2}^{w-1} Pr(Y_{i,t-(w-r)} = j_r | Y_{i,t-(w-r+1)} = j_{r-1}) \right\} \\ & \times Pr(Y_{it} = 1 | Y_{i,t-1} = j_{w-1}), \end{aligned} \quad (7.188)$$

[Sutradhar (2008, Section 2)] where the conditional probabilities have the formulas as in (7.33).

7.3. (Section 7.6.1.2) [Limiting behavior of the adaptive design weights w_i]
Recall from (7.97) that the binary probability conditional on δ_i (treatment indicator for the ith patient) is given by

$$\pi^*_{it}(\delta_i) = E(Y_{it}|\delta_i, x^*_{it}) = \frac{\exp(\theta_{it})}{1+\exp(\theta_{it})},$$

where $\theta_{it} = x'_{it}\beta$, with $x_{it} = (\delta_i, x^{*'}_{it})'$. Suppose that as $i \to \infty$,

$$(1) \quad (1/iT)\sum_{r=1}^{i-T}\sum_{j=1}^{T} \pi^*_{rj}(1) \to \pi_1, \quad (2) \quad (1/iT)\sum_{r=1}^{i-T}\sum_{j=1}^{T} \pi^*_{rj}(0) \to \pi_2,$$

$$(3) \quad (1/iT)\sum_{r=1}^{i-T}\sum_{j=1}^{T} u_{rj} \to u^*.$$

Then prove that as $i \rightarrow \infty$, the w_i defined in (7.104) converges to ω^* given by

$$\begin{aligned}\omega^* &= (1/(G+\tau))\left[(G-u^*+\pi_1\tau)\omega^* + (u^*+(1-\pi_2)\tau)(1-\omega^*)\right] \\ &= \{u^*+(1-\pi_2)\tau\} / \{2u^*+(2-\pi_1-\pi_2)\tau\},\end{aligned} \tag{7.189}$$

which is primarily a function of τ. Also argue that this ω^* is the limiting value of the probability of allocation of treatment A.

7.4. (Section 7.7.1.2) [Moment estimation of ρ for LBDR model]
For $\ell = 1$, it follows from (7.129) that

$$\rho_1 = \frac{\mu_{i,t-1}\{\rho+(1-\rho)\pi_{it}-\mu_{it}\}}{[\sigma_{itt}\ \sigma_{i,t-1,t-1}]^{1/2}}, \text{ for all } i = 1,\ldots,K,$$

where μ_{it} and σ_{itt} by (7.127) and (7.128), have the forms

$$\begin{aligned}\mu_{it} &= \sum_{j=1}^{t-1}(\pi_{ij}-\pi_{i,j+1})\rho^{t-j} + \pi_{it} \text{ and} \\ \text{var}[Y_{it}] &= \sigma_{itt} = \mu_{it}(1-\mu_{it}),\end{aligned}$$

respectively, which are functions of ρ for given β. Next, consider the sample lag 1 correlation

$$r_1 = \frac{\sum_{i=1}^{K}\sum_{t=2}^{T}\left(\frac{y_{it}-\mu_{it}}{\sqrt{\sigma_{itt}}}\right)\left(\frac{y_{i,t-1}-\mu_{i,t-1}}{\sqrt{\sigma_{i,t-1,t-1}}}\right)}{\sum_{i=1}^{K}\sum_{t=1}^{T}\left(\frac{y_{it}-\mu_{it}}{\sqrt{\sigma_{itt}}}\right)^2}. \tag{7.190}$$

Now by treating all σ_{itt} as functions of known ρ, and also by treating μ_{it} in r_1 as functions of known ρ, justify that a moment estimate for ρ may be obtained by using the iterative equation

$$\hat{\rho}(k+1) = \hat{\rho}(k) - \left([f'(\rho)]^{-1}f(\rho)\right)_{(k)}, \tag{7.191}$$

where $(\cdot)_{(k)}$ indicates that the quantity in $(\cdot)$ is evaluated at $\rho = \hat{\rho}(k)$, with

$$\begin{aligned}f(\rho) &= r_1 - \rho_1 \\ f'(\rho) &\simeq -\sum_{i=1}^{K}\sum_{t=2}^{T}\frac{1}{\sqrt{\sigma_{itt}\sigma_{i,t-1,t-1}}}\left[\mu_{i,t-1}\{(1-\pi_{it}) - \frac{\partial\mu_{it}}{\partial\rho}\}\right. \\ &\left. + \frac{\partial\mu_{i,t-1}}{\partial\rho}\{\rho+(1-\rho)\pi_{it}-\mu_{it}\}\right].\end{aligned} \tag{7.192}$$

7.5. (Section 7.7.2.2.1) [First-order derivatives for GQL estimation]
Verify that the first-order derivatives, namely $\partial\mu_i'/\partial\zeta$ for (7.151) may be computed by using

$$\frac{\partial\mu_{it}}{\partial\beta_m} = \{\tilde{p}_{it}(1-\tilde{p}_{it})\mu_{i,t-1} + \pi_{it}(1-\pi_{it})(1-\mu_{i,t-1})\}x_{itm}$$

$$+(\tilde{p}_{it} - \pi_{it})\frac{\partial \mu_{i,t-1}}{\partial \beta_m}, \tag{7.193}$$

for $m = 1, \ldots, p$, and

$$\frac{\partial \mu_{it}}{\partial \theta} = \sum_{u=2}^{t}\left[\mu_{i,u-1}\tilde{p}_{iu}(1-\tilde{p}_{iu})\prod_{k=u+1}^{t}(\tilde{p}_{ik} - \pi_{ik})\right], \tag{7.194}$$

for $t = 2, \ldots, T$, and where $\partial \mu_{i1}/\partial \theta = 0$.

7.6. (Section 7.7.2.2.2) [Basic sufficient statistics for OGQL estimation]
Recall from (7.142) that the marginal probability of y_{i1} is p_{i1}, whereas $p_{it|t-1}$ for $t = 2, \ldots, T$ denote the lag 1 conditional probabilities. This yields the likelihood of the data as

$$L(\beta, \theta) = \prod_{i=1}^{K} f(y_{i1}) f(y_{i2}|y_{i1}) \ldots f(y_{iT}|y_{i,T-1}), \tag{7.195}$$

with

$$f(y_{i1}) = \mu_{i1}{}^{y_{i1}}[1-\mu_{i1}]^{1-y_{i1}} \text{ and } f(y_{it}|y_{i,t-1}) = (p_{it|t-1})^{y_{it}}(1-p_{it|t-1})^{1-y_{it}},$$

where by (7.142),

$$\mu_{i1} = p_{i1} = \pi_{i1} = \frac{\exp(x_{i1}'\beta)}{1+\exp(x_{i1}'\beta)},$$

and

$$p_{it|t-1} = \frac{\exp(x_{it}'\beta + \theta y_{i,t-1})}{1+\exp(x_{it}'\beta + \theta y_{i,t-1})} \text{ for } t = 2, \ldots, T.$$

Simplify the likelihood function $L(\beta, \theta)$ as

$$L(\beta, \theta) = \prod_{i=1}^{K}\left[\frac{\exp[(x_{i1}'\beta)y_{i1}]}{1+\exp(x_{i1}'\beta)}\prod_{t=2}^{T}\frac{\exp[(x_{it}'\beta + \theta y_{i,t-1})y_{it}]}{1+\exp(x_{it}'\beta + \theta y_{i,t-1})}\right],$$

and argue that the pairwise products $y_{i1}y_{i2}, \ldots, y_{iu}y_{it}, \ldots, y_{i,T-1}y_{iT}$ along with the first-order responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ provide sufficient information for the estimation of β and θ.

References

1. Amemiya, T. (1985). *Advanced Econometrics,* Cambridge, MA: Harvard University Press.
2. Bahadur, R.R. (1961). A representation of the joint distribution of responses to *n* dichotomous items. In *Studies in Item Analysis and Prediction. Stanford Mathematical Studies in the Social Sciences*; Solomon, H., Ed.; Vol. 6, 158 – 168.
3. Bandyopadhyay, U. & Biswas, A. (1999). Allocation by randomized play-the-winner rule in the presence of prognostic factors. *Sankhya A* 61, 396 – 412.
4. Cox, D. R. (1972). The analysis of multivariate binary data. *Appl. Statist.,* 21, 113 – 120.

5. Cox, D. R. & Lewis, P. W. (1966). *The Statistical Analysis of Series of Events.* London: Chapman & Hall.
6. Farrell, P. J. & Sutradhar, B. C. (2006). A non-linear conditional probability model for generating correlated binary data. *Statist. Probab. Lett.,* 76, 353 – 361.
7. Jacobs, P. A. & Lewis, P. A. W. (1983). Stationary discrete autoregressive moving average generated by mixtures. *J. Time Series. Anal.,* 4, 19 – 36.
8. Kanter, M. (1975). Autoregression for discrete processes mod 2. *J. Appl. Probab.*, 12, 371 – 375.
9. Keenan, D. M. (1982). A time series analysis of binary data. *J. Am. Stat. Assoc.,* 77, 816 – 821.
10. Klotz, J. (1973). Statistical inference in Bernoulli trials with dependence. *Ann. Statist.,* 1, 373 – 379.
11. Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika,* 73, 13 – 22.
12. Lindquist, B. (1978). A note on Bernoulli trials with dependence. *Scand. J. Statist.,* 5, 205 – 208.
13. Manski, C. F. (1987). Semi-parametric analysis of random effects linear models from binary panel data. *Econometrica,* **55,** 357 – 362.
14. Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.
15. Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observations. *Biometrics,* 44, 1033 – 1048.
16. Qaqish, B.F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika*, 90, 455 – 463.
17. Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statist. Sci.,* **18**, 377 – 393.
18. Sutradhar, B. C. (2008). On auto-regression type dynamic mixed models for binary panel data. *Metron,* 66, 205 – 217.
19. Sutradhar, B. C. (2010). Inferences in generalized linear longitudinal mixed models. *The Canad. J. Statist.,* 38, 174 – 196.
20. Sutradhar, B.C ., Biswas, A., & Bari, W. (2005). Marginal regression for binary longitudinal data in adaptive clinical trials. *Scand. J. Statist.,* **32,** 93 – 113.
21. Sutradhar, B. C. & Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika,* 86, 459 – 65.
22. Sutradhar, B. C. & Farrell, P. J. (2007). On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data. *Sankhya B,* 69, 448 – 467.
23. Sutradhar, B. C. & Jowaheer, V. (2010). Treatment design selection effects on parameter estimation in dynamic logistic models for longitudinal binary data. *J. Statist. Comput. Simul.,* 80, 1053 – 1067.
24. Sutradhar, B.C. & Kovacevic, M. (2000). Analysing Ordinal Longitudinal Survey Data : Generalized Estimating Equations Approach. *Biometrika*, **87**, 837 – 848.
25. Sutradhar, B. C., Rao, R. P,& Pandit, V. N. (2008). Generalized method of moments versus generalized quasi-likelihood inferences in binary panel data models. *Sankhya B,* 70, 34-62.
26. Tagore, V. and Sutradhar, B. C. (2009). Conditional inference in linear versus nonlinear models for binary time series. *J. Statist. Comput. Simul.,* 79, 881 – 897.
27. Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.
28. Wei, L. J. & Durham, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.,* 73, 840 – 843.
29. Zeger, S.L., Liang, K.Y., & Albert, P.S. (1998). Models for longitudinal data: A generalized estimating equations approach. *Biometrics*, 44, 1049 – 1060.
30. Zeger, S. L., Liang, K-Y., & Self, S. G. (1985). The analysis of binary longitudinal data with time independent covariates. *Biometrika,* 72, 31 – 38.
31. Zhao, L. P. & Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika,* 77, 642 – 648.

Appendix

SLID Data: Tables 7A to 7E
Asthma Data: Table 7F

Table 7A. Estimated counts cross-classified according to ‘Unemployment Status’ and ‘Sex’ (in ’000).

		Year			
Sex	Unemployment Status	1993	1994	1995	1996
Male	Not unemployed (=0)	7357	7344	7361	7346
	Unemployed (=1)	139	152	135	150
Female	Not unemployed (=0)	8094	8029	8045	8060
	Unemployed (=1)	141	206	190	175

Table 7B. Estimated counts cross-classified according to ‘Age Group in 1993’ and ‘Unemployed All Year’ (in ’000).

		Year			
Age Group	Unemployment Status	1993	1994	1995	1996
$16 \leq$ Age in 1993 <25	Not unemployed (=0)	2319	2299	2300	2299
	Unemployed (=1)	34	54	53	54
$25 \leq$ Age in 1993 < 55	Not unemployed (=0)	10978	10917	10938	10397
	Unemployed (=1)	198	259	238	239
$55 \leq$ Age in 1993 < 65	Not unemployed (=0)	2154	2157	2168	2170
	Unemployed (=1)	48	45	34	32

Table 7C. Estimated counts cross-classified by ‘Region of Residence in 1993’ and ‘Unemployed All Year’ (in ’000).

		Year			
Region of Residence	Unemployment Status	1993	1994	1995	1996
Atlantic	Not unemployed (=0)	3472	3424	3424	3385
	Unemployed (=1)	90	112	102	119
Quebec	Not unemployed (=0)	3244	3216	3212	3228
	Unemployed (=1)	80	110	112	97
Ontario	Not unemployed (=0)	3787	3793	3822	3820
	Unemployed (=1)	64	69	52	68
Prairies	Not unemployed (=0)	3584	3554	3555	3569
	Unemployed (=1)	33	51	44	29
BC	Not unemployed (=0)	1364	1386	1393	1404
	Unemployed (=1)	13	16	15	12

Table 7D. Estimated counts cross-classified according to ‘Education Level’ and ‘Unemployed All Year’ (in ’000).

		Year			
Education Level	Unemployment Status	1993	1994	1995	1996
Low education	Not unemployed (=0)	3244	2990	2908	2872
	Unemployed (=1)	115	122	122	111
Medium education	Not unemployed (=0)	10165	10252	10274	10241
	Unemployed (=1)	154	215	188	198
High education	Not unemployed (=0)	2042	2131	2224	2293
	Unemployed (=1)	11	21	15	16

Table 7E. Estimated counts cross-classified by ‘Marital Status’ and ‘Unemployed All Year’ (in ’000).

		Year			
Marital Status	Unemployment Status	1993	1994	1995	1996
Married/common law	Not unemployed (=0)	10832	10853	10973	11109
	Unemployed (=1)	175	225	177	180
Separated/divorced	Not unemployed (=0)	1008	1107	1192	1289
	Unemployed (=1)	30	34	56	45
Widowed	Not unemployed (=0)	300	332	369	391
	Unemployed (=1)	5	6	7	10
Single	Not unemployed (=0)	3311	3081	2872	2617
	Unemployed (=1)	70	93	85	90

Table 7F. Asthma data for 537 children from Steubenville, Ohio, from ages 7 through 10.

		Ages (Occasions)			
Covariates/Response	Child Identity	7(1)	8(2)	9(3)	10(4)
Intercept	1 to 537	1	1	1	1
Mother's smoking status	1 to 350	0	0	0	0
	351 to 537	1	1	1	1
Asthma status	1 to 237	0	0	0	0
	238 to 247	0	0	0	1
	248 to 262	0	0	1	0
	263 to 266	0	0	1	1
	267 to 282	0	1	0	0
	283 to 284	0	1	0	1
	285 to 291	0	1	1	0
	292 to 294	0	1	1	1
	295 to 318	1	0	0	0
	319 to 321	1	0	0	1
	322 to 324	1	0	1	0
	325 to 326	1	0	1	1
	327 to 332	1	1	0	0
	333 to 334	1	1	0	1
	335 to 339	1	1	1	0
	340 to 350	1	1	1	1
	351 to 468	0	0	0	0
	469 to 474	0	0	0	1
	475 to 482	0	0	1	0
	483 to 484	0	0	1	1
	485 to 495	0	1	0	0
	496	0	1	0	1
	497 to 502	0	1	1	0
	503 to 506	0	1	1	1
	507 to 513	1	0	0	0
	514 to 516	1	0	0	1
	517 to 519	1	0	1	0
	520	1	0	1	1
	521 to 524	1	1	0	0
	525 to 526	1	1	0	1
	527 to 530	1	1	1	0
	531 to 537	1	1	1	1

Chapter 8
Longitudinal Mixed Models for Count Data

Recall that in Chapter 6, a class of correlation models was discussed for the analysis of longitudinal count data collected from a large number of independent individuals, whereas in Chapter 4, we discussed the analysis of count data collected from the members of a large number of independent families. Thus, in Chapter 4, familial correlations among the responses of the members of a given family were assumed to be caused by the influence of the same family effect on the members of the family, whereas in Chapter 6, longitudinal correlations were assumed to be generated through a dynamic relationship among the repeated counts collected from the same individual. A comparison between the models in these two chapters (4 and 6) clearly indicates that modelling the longitudinal correlations for count data through a common individual random effect would be inappropriate. If it is, however, thought that the longitudinal count responses may also be influenced by an invisible random effect due to the individual, this will naturally create a complex correlation structure where repeated responses will satisfy a longitudinal correlation structure but conditional on the individual random effect. The purpose of this chapter is to discuss inferences in such longitudinal mixed models that generate longitudinal correlations conditional on the individual random effect. Note that this type of longitudinal mixed models is studied by some econometricians among others, where the model is referred to as the panel data model for count data. For example, we refer to Hausman, Hall and Griliches (1984), Wooldridge (1999), and Montalvo (1997). See also Sutradhar and Bari (2007), and Jowaheer and Sutradhar (2009). Further note that the longitudinal mixed model for count data discussed in this chapter is extended in Chapter 10 to the familial longitudinal mixed model.

8.1 A Conditional Serially Correlated Model

Let $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ be the T repeated counts collected from the ith $(i = 1, \ldots, K)$ individual, $x_{it} = (x_{it1}, \ldots, x_{itj}, \ldots, x_{itp})'$ be the p-dimensional covariate vector associated with the response y_{it}, and $\beta = (\beta_1, \ldots, \beta_j, \ldots, \beta_p)'$ denote the regression

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_8,

effects of x_{it} on y_{it}. Because the repeated responses are likely to be correlated, in Chapter 6, the regression effect β was estimated by taking the correlations of the repeated data into account. More specifically, a class of autocorrelations was introduced and the autocorrelation parameters involved in such a correlation structure were consistently estimated in order to obtain a consistent and efficient estimate for the regression effect β. The generalized quasi-likelihood (GQL) method was used for such efficient estimation.

In this chapter we assume that the repeated counts of an individual are also influenced by the individual random effect. Thus, conditional on the random effect $\gamma_i \overset{i.i.d.}{\sim} N(0, \sigma_\gamma^2)$, the repeated responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are assumed to follow a suitable autocorrelation structure such as the class of autocorrelations discussed in Chapter 6. However, for simplicity the inference in this chapter is given under a conditional AR(1) correlation structure only. To be specific, conditional on γ_i, let $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ follow the dynamic relationship as in (6.44), that is,

$$y_{it}|\gamma_i = \rho * y_{i,t-1}|\gamma_i + d_{it}|\gamma_i, \;\; t = 2, \ldots, T, \tag{8.1}$$

where it is assumed that $y_{i,1}|\gamma_i \sim \text{Poi}(\mu_{i1}^*)$, and for $t = 2, \ldots, T$, $y_{i,t-1}|\gamma_i \sim \text{Poi}(\mu_{i,t-1}^*)$ and $d_{it}|\gamma_i \sim \text{Poi}(\mu_{it}^* - \rho\mu_{i,t-1}^*)$ with $\mu_{ij}^* = \exp(x_{ij}'\beta + \gamma_i)$ for $j = 1, \ldots, t-1, t, \ldots, T$. In (8.1), conditional on γ_i, d_{it} and $y_{i,t-1}$ are independent. Furthermore, for a given count $y_{i,t-1}$, $\rho * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\rho)$ is a binomial thinning operation as defined in Section 6.3 [see also (6.44)]. Here $b_j(\rho)$ stands for a binary variable with $Pr[b_j(\rho) = 1] = \rho$ and $Pr[b_j(\rho) = 0] = 1 - \rho$. It then follows that the mean and the variance of y_{it} conditional on γ_i are given by

$$E(Y_{it}|\gamma_i) = \text{var}(y_{it}|\gamma_i) = \mu_{it}^* = \exp(x_{it}'\beta + \gamma_i). \tag{8.2}$$

Furthermore, by using (8.1), for $u < t$, one can compute the $E(Y_{iu}Y_{it}|\gamma_i)$ which yields the lag $(t-u)$ correlation conditional on γ_i as

$$\text{corr}(Y_{iu}, Y_{it}|\gamma_i) = \rho^{t-u}\sqrt{\frac{\mu_{iu}^*}{\mu_{it}^*}}, \tag{8.3}$$

where $\mu_{it}^* = \exp(x_{it}'\beta + \gamma_i)$. Note that the conditional serial correlation given by (8.3) under the non-stationary model (8.1) depends on the time-dependent covariates, similar to (6.46). This conditional correlation in (8.3), does not, however, depend on γ_i, and it is clear, based on the positive parameter of the Poisson distribution of d_{it}, that ρ must now satisfy the range restriction $0 < \rho < \min[1, \mu_{it}^*/\mu_{i,t-1}^*]$, which is the same as

$$0 < \rho < \min[1, m_{it}/m_{i,t-1}] \;\; \text{for } t = 2, \ldots, T; i = 1, \ldots, K, \tag{8.4}$$

where $m_{it} = \exp(x_{it}'\beta)$.

8.1.1 Unconditional Mean, Variance, and Correlations Under Serially Correlated Model

Similar to Chapter 4, we assume that $\gamma_i \overset{i.i.d.}{\sim} N(0, \sigma_\gamma^2)$. It then follows from (8.2) that y_{it} unconditionally has the mean and the variance given by

$$E[Y_{it}] = \exp(x'_{it}\beta + \sigma_\gamma^2/2) = \mu_{it}, (say), \quad (8.5)$$

$$\text{var}[Y_{it}] = \mu_{it} + [\exp(\sigma_\gamma^2) - 1]\mu_{it}^2 = \sigma_{itt}. \quad (8.6)$$

Now to derive the unconditional covariance between y_{iu} and y_{it}, it follows from the model (8.1) that conditional on γ_i, the covariance between y_{iu} and y_{it} $(u < t)$ is given by

$$\text{cov}(Y_{iu}, Y_{it}|\gamma_i) = \rho^{t-u}\mu_{iu}^*. \quad (8.7)$$

Consequently, the unconditional covariance between y_{iu} and y_{it} has the form

$$\begin{aligned} \sigma_{iut} &= \text{cov}(Y_{iu}, Y_{it}) \\ &= E[\text{cov}\{(Y_{iu}, Y_{it})|\gamma_i\}] + \text{cov}[E(Y_{iu}|\gamma_i), E(Y_{it}|\gamma_i)] \\ &= \rho^{t-u}\mu_{iu} + [\exp(\sigma_\gamma^2) - 1]\mu_{iu}\mu_{it}, \end{aligned} \quad (8.8)$$

leading to the lag $t-u$ correlations as

$$\text{corr}(Y_{iu}, Y_{it}) = \frac{\mu_{iu}\rho^{t-u} + \mu_{iu}\mu_{it}\{\exp(\sigma_\gamma^2) - 1\}}{[\{\mu_{iu} + (\exp(\sigma_\gamma^2) - 1)\mu_{iu}^2\}\{\mu_{it} + (\exp(\sigma_\gamma^2) - 1)\mu_{it}^2\}]^{1/2}}. \quad (8.9)$$

Note that the unconditional mean and the variance given by (8.5) and (8.6), respectively, have the same form as in (4.5) and (4.6) under the familial model, but the unconditional correlation given in (4.8) under the familial model (see Chapter 4) is different from the unconditional correlation (8.9) under the present longitudinal mixed model (8.1). More specifically, the correlation between any two members in a familial setup is computed under the assumption that the responses of the members are independent conditional on the random family effect, whereas the correlation in (8.9) is computed under the assumption that two repeated responses of an individual are correlated conditional on the individual random effect, ρ being the correlation index parameter.

8.2 Parameter Estimation

Recall that it was found in Chapter 6 that in a longitudinal setup, one may consistently and efficiently estimate the regression effects β by solving the GQL estimating equation (6.56), whereas the correlation index parameter ρ was consistently estimated by using the moment equation (6.58). In the present longitudinal mixed model setup (8.1), these two parameters may still be computed by using similar equations, but we need to compute an additional parameter, namely σ^2_γ, the variance of the random effects. In the familial set up, that is, in Chapter 4 we have computed the regression effects β and the variance of the random family effects σ^2_γ by solving the GQL estimating equations (4.62) and (4.69), respectively. In the following sections, we use this GQL approach for the estimation of all three parameters β, σ^2_γ, and ρ. Also, recall from Chapter 3 (see Section 3.3.1) that a generalized method of moments (GMM) [or an improved method of moments (IMM)] as well as the GQL approaches were used to estimate the parameters of a dynamic linear mixed model. Jowaheer and Sutradhar (2009) have compared the efficiencies of these GMM and GQL approaches for the panel count data. In the following sections, in addition to the GQL approach, we also present the GMM approach for the estimation of all three parameters. Furthermore, we discuss a conditional maximum likelihood (CML) approach due to Wooldridge (1999), only for β estimation. For the estimation of the same β parameter, in Section 8.2.1, we discuss an instrumental variables based generalized method of moments (IVBGMM) studied by Montalvo (1997), among others.

8.2.1 *Estimation of the Regression Effects* β

8.2.1.1 GMM/IMM Approach

Note that when the traditional method of moments (MM) is used to estimate the β vector, one solves the unbiased moment estimating equation

$$\sum_{i=1}^{K} \psi_{i1}(\beta,\, \sigma^2_\gamma) = 0, \tag{8.10}$$

where

$$\psi_{1i}(\beta,\, \sigma^2_\gamma) = \sum_{t=1}^{T} \left[x_{it}(y_{it} - \mu_{it})\right], \tag{8.11}$$

[Jiang (1998); Sutradhar (2004)] is an unbiased moment function as $E[Y_{it}] = \mu_{it}$ leading to $E[\psi_{1i}(\beta,\, \sigma^2_\gamma)] = 0$. In the notation of $(8.10) - (8.11)$ but for known σ^2_γ, in the GMM approach [Hansen (1982)] one would minimize the distance function

$$Q(\beta) = K^{-1}\left[\sum_{i=1}^{K} \psi_{i1}(\beta|\sigma^2_\gamma)\right]' C_1 \left[\sum_{i=1}^{K} \psi_{i1}(\beta|\sigma^2_\gamma)\right], \tag{8.12}$$

where C_1 is a suitable weight. An optimal choice for C_1 would be the inverse of the variance of the unbiased moment function; that is,

$$C_1 = [\text{var}(K^{-1}\sum_{i=1}^{K}\psi_{i1}(\beta,\sigma_\gamma^2))]^{-1}. \tag{8.13}$$

Note that minimizing $Q(\beta)$ in (8.12) with respect to β is equivalent to solving the GMM estimating equation

$$\frac{\partial \psi_1'}{\partial \beta} C_1 \psi_1 = 0, \tag{8.14}$$

for β, where $\psi_1 = K^{-1}\sum_{i=1}^{K}\psi_{1i}$ so that

$$\frac{\partial \psi_1'}{\partial \beta} = K^{-1}\sum_{i=1}^{K}\sum_{t=1}^{T}\mu_{it}x_{it}x_{it}'. \tag{8.15}$$

As far as the C_1 matrix in (8.14) is concerned, by (8.13), this may be computed by

$$C_1^{-1} = K^{-2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{t=1}^{T}\sigma_{iut}(\beta,\ \sigma_\gamma^2,\ \rho)x_{iu}x_{it}', \tag{8.16}$$

where for all $u,t = 1,\ldots,T$, $\sigma_{iut}(\beta,\ \sigma_\gamma^2,\ \rho)$ is given by (8.6) and (8.8).

8.2.1.2 GQL Approach

For $y_i = [y_{i1},\ldots,y_{it},\ldots,y_{iT}]'$, let

$$E[Y_i] = \mu_i(\beta,\sigma^2) = [\mu_{i1}(\beta,\sigma^2),\ldots,\mu_{it}(\beta,\sigma^2),\ldots,\mu_{iT}(\beta,\sigma^2)]',$$

with $\mu_{it}(\beta,\sigma_\gamma^2) = \exp(x_{it}'\beta + \sigma_\gamma^2)$ as given in (8.5). Also, let $\Sigma_i(\beta,\sigma^2,\rho)$ denote the covariance matrix of y_i. To be specific,

$$\Sigma_i(\beta,\sigma_\gamma^2,\rho) = (\sigma_{iut}), \tag{8.17}$$

where $\text{var}(Y_{it}) = \sigma_{itt} \equiv \sigma_{itt}(\beta,\sigma_\gamma^2)$ and $\text{cov}(Y_{iu},Y_{it}) = \sigma_{iut} \equiv \sigma_{iut}(\beta,\sigma_\gamma^2,\rho)$ for $u \neq t$, with σ_{itt} and σ_{iut} defined as in (8.6) and (8.8), respectively. Now following (4.62) or (6.56) [see also Sutradhar (2004), and Sutradhar and Jowaheer (2003)], one may solve the generalized quasi-likelihood estimating equation given by

$$\sum_{i=1}^{K}\frac{\partial \mu_i'(\beta,\sigma_\gamma^2)}{\partial \beta}\Sigma_i^{-1}(\beta,\sigma_\gamma^2,\rho)[y_i - \mu_i(\beta,\sigma_\gamma^2)] = 0, \tag{8.18}$$

to obtain the GQL estimate of β. For given σ_γ^2 and ρ, the GQL estimate obtained from (8.18) is consistent for β. This is because, as $E(Y_i) = \mu_i(\beta,\sigma^2)$, the estimating equation (8.18) is unbiased. Furthermore, because the GQL estimating equation (8.18) is constructed by using the covariance matrix $\Sigma_i(\beta,\sigma^2,\rho)$ as a weight matrix, it follows that the GQL estimate of β obtained from (8.18) would be highly efficient as compared to other competitors such as the GMM estimator.

8.2.1.3 Conditional Maximum Likelihood (CML) Approach

Note that as opposed to the correlation parameters (σ_γ^2, ρ) based moment (8.14) and GQL (8.18) estimation for β, some authors such as Hausman, Hall, and Griliches (1984), Montalvo (1997), and Wooldridge (1999) have used a 'working' independence, that is, $\rho = 0$ assumption based likelihood estimation. Note that at a given point of time, conditional on γ_i, it may be assumed that the count response y_{it} follows the Poisson distribution

$$f_{it}(y_{it}|\gamma_i) = \frac{\exp(-\mu_{it}^*)\mu_{it}^{*y_{it}}}{y_{it}!}, \tag{8.19}$$

with $\mu_{it}^* = \exp(x_{it}'\beta + \gamma_i)$. When it is assumed that the repeated responses $y_{i1}, \ldots, y_{iT}$ are independent conditional on the individual random effect γ_i, one may write a 'working' joint density as

$$L_i(\beta, \gamma_i) = \Pi_{i=1}^T f_{it}(y_{it}|\gamma_i) = \frac{\exp(-\sum_{t=1}^T \mu_{it}^*)\Pi_{t=1}^T \mu_{it}^{*y_{it}}}{\Pi_{t=1}^T y_{it}!}. \tag{8.20}$$

Note that this 'working' likelihood model ignoring the serial dependence is capable of producing the correct mean and the variance, provided the distribution of γ_i is known. This is, however, well known (see also Chapter 4) that even if it is assumed that $\gamma_i \stackrel{i.i.d.}{\sim} N(0, \sigma_\gamma^2)$, the exact likelihood estimation of β and σ_γ^2 is complicated. To avoid such complexity, some authors such as Wooldridge [1999, eqn. (2.6), p. 79] used a further conditioning on the total count of an individual and proposed a conditional maximum likelihood approach for the estimation of β. This approach is not influenced by the individual random effects and hence they may follow any distributions. We first explain this approach below and then discuss its limitations.

In the CML approach, conditional on total count $\sum_{t=1}^T y_{it} = n_i$, one first writes a conditional likelihood for the repeated responses under the ith individual as

$$\begin{aligned} L_i(\beta|n_i) &= f_i(y_{i1}, \ldots, y_{iT}|n_i) \\ &= \frac{n_i!}{y_{i1}! \ldots y_{i,T-1}!(n_i - \sum_{t=1}^{T-1} y_{it})!} p_{i1}^{y_{i1}} \cdots p_{i,T-1}^{y_{i,T-1}} p_{iT}^{n_i - \sum_{t=1}^{T-1} y_{it}}, \end{aligned} \tag{8.21}$$

where $p_{it} = \mu_{it}^* / \sum_{t=1}^T \mu_{it}^*$. Note that because

$$p_{it} = \frac{\exp(x_{it}'\beta + \gamma_i)}{\exp(\gamma_i)\sum_{t=1}^T \exp(x_{it}'\beta)} = \frac{\exp(x_{it}'\beta)}{\sum_{t=1}^T \exp(x_{it}'\beta)}, \tag{8.22}$$

the conditional likelihood (8.21) is free from γ_i. Consequently, one may estimate β [Montalvo (1997, Section 1); Wooldridge (1999, eqn. (2.6), p. 79)] by maximizing the log-likelihood

$$
\begin{aligned}
L^*(\beta) &= \Pi_{i=1}^{K} \log L_i(\beta|n_i) \\
&= k_0 + \sum_{i=1}^{K}\sum_{t=1}^{T} y_{it} \log(p_{it}),
\end{aligned} \tag{8.23}
$$

where k_0 is a constant free from β, and $y_{iT} = n_i - \sum_{t=1}^{T-1} y_{it}$.

Note that the maximization of the log-likelihood function (8.23) for β is equivalent to solving the likelihood estimating equation given by

$$
\frac{\partial L^*(\beta)}{\partial \beta} = \sum_{i=1}^{K}\sum_{t=1}^{T} y_{it}\left[x_{it} - \sum_{t=1}^{T} p_{it}x_{it}\right] = 0. \tag{8.24}
$$

Further note that the CML estimate of β obtained from (8.24) is expected to be consistent but inefficient. The inefficiency arises mainly because of conditioning on the cluster total as well as for ignoring the serial correlations. More specifically, when the data are serially correlated, the independence assumption based conditional likelihood (8.20) is no longer a valid likelihood. Hence it is bound to produce an inefficient estimate. A simulation study in Section 8.2.1.5 for the estimation of β also supports this observation. More specifically it is shown that when count panel data are generated by using the AR(1) type mixed model (8.1), the CML estimates for the components of β are unbiased and hence consistent, but these estimates are less efficient than the GQL approach given in the previous section, where the GQL estimation of β also utilizes the information about other parameters of the longitudinal mixed model.

There appears to be another major problem with the CML approach. When the covariates are stationary such as $x_{it} = \tilde{x}_i$ for $t = 1,\ldots,T$, the multinomial probability in (8.22) reduces to $p_{it} = 1/T$ which makes the likelihood function (8.21) parameter free. Thus in the stationary case β becomes redundant and there is nothing to estimate. Hence the inference procedure breaks down. In the GQL approach, one may still estimate β consistently and efficiently, even if the data are stationary.

8.2.1.4 Instrumental Variables Based GMM (IVBGMM) Estimation Approach

As opposed to the conditioning on the total count as in the last section, Montalvo (1997, eqn. 32, p.85) considered the lag 1 based differences, namely

$$
\psi_{it}(\beta) = y_{it} - y_{i,t-1}\exp[(x_{it} - x_{i,t-1})'\beta], \;\; \text{for } t = 2,\ldots,T, \tag{8.25}
$$

which is unbiased for zero irrespective of the distribution of γ_i. Next by exploiting the $(T-1)\times 1$ vector

$$
\psi_i(\beta) = [\psi_{i2}(\beta),\ldots,\psi_{it}(\beta),\ldots,\psi_{iT}(\beta)]',
$$

Montalvo (1997, eqn. 36) obtained a GMM estimate for β by minimizing the quadratic distance function

$$D = \left[\sum_{i=1}^{K} Z_i' \psi_i(\beta)\right]' W^{-1} \left[\sum_{i=1}^{K} Z_i' \psi_i(\beta)\right], \tag{8.26}$$

where Z_i is the $(T-1) \times p\{(T(T+1))/2-1\}$ instrumental matrix given by

$$Z_i = \begin{bmatrix} z_{i2} & 0 & 0 \cdots & 0 \\ 0 & z_{i3} & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \cdots & z_{iT} \end{bmatrix}, \tag{8.27}$$

with $z_{it} = [x_{it}', x_{i(t-1)}', \ldots, x_{i1}']$, and where

$$W = \frac{1}{K} \sum_{i=1}^{K} Z_i' \psi_i(\beta) \psi_i'(\beta) Z_i.$$

Note that obtaining β by minimizing the distance function D in (8.26) is equivalent to solving the estimating equation

$$\left[\sum_{i=1}^{K} \frac{\partial \psi_i'}{\partial \beta} Z_i\right] W^{-1} \left[\sum_{i=1}^{K} Z_i' \psi_i(\beta)\right] = 0, \tag{8.28}$$

for β, where $\partial \psi_i' / \partial \beta$ is obtained by using the formula for the general element

$$\frac{\partial \psi_{it}}{\partial \beta} = -y_{i,t-1}[x_{it} - x_{i,t-1}] \exp((x_{it} - x_{i,t-1})'\beta.$$

But the use of a sandwich-type covariance matrix estimate $\hat{W}(\hat{\beta}_r)$ in the distance function D may cause bias, and hence inconsistency, because of the repeated use of iterative estimated values for the parameter of interest. Furthermore, the IVBGMM estimate obtained from (8.28) will in fact produce a less efficient estimate than the CML approach. This is because the IVBGMM estimating equation (8.28) uses only lag 1 pairwise responses, whereas the CML approach uses all possible responses in the cluster to form the likelihood function. Consequently, the IVBGMM estimates will be much more inefficient than the GQL estimates, as the GQL estimates are expected to be more efficient than the CML estimates. The empirical study in the following section also supports this conjecture. We also refer to Jowaheer and Sutradhar (2009), where it is shown that the GMM approach given in Section 8.2.1.1 produces asymptotically less efficient estimates than the generalized quasi-likelihood approach.

Note that similar to the CML approach, the IVBGMM approach becomes useless in the stationary case. This is because when $x_{it} = \tilde{x}_i$, for all $t = 1, \ldots, T$, the lag 1 based difference $\psi_{it}(\beta)$ in (8.25) does not contain any parameter. Thus, there is nothing to estimate by minimizing the distance function (8.26), and the inference procedure breaks down.

8.2.1.5 A Simulation Study

The CML and IVBGMM approaches are introduced for the estimation of β only, thus in this section, we make a mean squared error based efficiency comparison of these approaches with the GQL approach for the estimation of β. This is done through a simulation study. The GQL approach appears to be the best among these three approaches, the IVBGMM approach being the worst. In Section 8.2.2.3, we provide an asymptotic efficiency comparison between the GQL and GMM/IMM approaches for the estimation of both main parameters β and σ^2_γ.

Let $\hat{\beta}_{GQL}$ be the solution of (8.18). For known σ^2_γ and ρ, this GQL estimate may be obtained by using the iterative equation

$$\hat{\beta}_{GQL}(r+1) = \hat{\beta}_{GQL}(r) + \left[\left\{\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1}(\rho,\, \sigma^2_\gamma) \frac{\partial \mu_i}{\partial \beta'}\right\}^{-1} \times \left\{\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1}(\rho,\, \sigma^2_\gamma)(y_i - \mu_i)\right\}\right]_{|\beta=\hat{\beta}_{GQL}(r)}, \quad (8.29)$$

where $\hat{\beta}_{GQL}(r)$ is the value of β at the rth iteration.

Note that the CML estimate of β is obtained by solving the multinomial likelihood estimating equation (8.24). Let $\hat{\beta}_{CML}$ denote this estimate, which we obtain by using the iterative equation

$$\hat{\beta}_{CML}(r+1) = \hat{\beta}_{CML}(r) - \left[\left\{\frac{\partial^2 L^*(\beta)}{\partial \beta \partial \beta'}\right\}^{-1} \frac{\partial L^*(\beta)}{\partial \beta}\right]_{|\beta=\hat{\beta}_{CML}(r)}$$

$$= \hat{\beta}_{CML}(r) + \left[\left\{\sum_{i=1}^{K}(\sum_{t=1}^{T} y_{it})\left(\sum_{t=1}^{T} p_{it}x_{it}x_{it}' - \sum_{t=1}^{T} p_{it}x_{it} \sum_{t=1}^{T} p_{it}x_{it}'\right)\right\}^{-1} \times \sum_{i=1}^{K}\sum_{t=1}^{T} y_{it}\left\{x_{it} - \sum_{t=1}^{T} p_{it}x_{it}\right\}\right]_{|\beta=\hat{\beta}_{ML}(r)}, \quad (8.30)$$

where $\hat{\beta}_{CML}(r)$ is the value of β at the rth iteration. Similarly, the IVBGMM estimate of β, say $\hat{\beta}_{IVBGMM}$, is obtained by solving the estimating equation (2.28). This may be achieved by using the iterative equation

$$\hat{\beta}_{IVBGMM}(r+1) = \hat{\beta}_{IVBGMM}(r) - \left[\left[\left\{\sum_{i=1}^{K} \frac{\partial \psi_i'}{\partial \beta} Z_i\right\} W^{-1} \left\{\sum_{i=1}^{K} Z_i' \frac{\partial \psi_i(\beta)}{\partial \beta'}\right\}\right]^{-1}\right.$$

$$\times \left[\sum_{i=1}^{K} \frac{\partial \psi_i'}{\partial \beta} Z_i\right] W^{-1} \left[\sum_{i=1}^{K} Z_i' \psi_i(\beta)\right]\Bigg]\Bigg|_{\beta=\hat{\beta}_{IVBGMM}(r)}, \quad (8.31)$$

where $\hat{\beta}_{IVBGMM}(r)$ is the value of β at the rth iteration.

For selected values of $\sigma_\gamma^2 = 0.5,\ 1.0$, and $\rho = 0.5$, we simulate the count data y_{it} by using the dynamic mixed model (8.1), for $t = 1,\ldots,T$, and $i = 1,\ldots,K$. We choose $K = 100$ independent individuals and generate data for them for a period of $T = 4$ time points. We consider $p = 2$ regression parameters and choose to estimate their true values $\beta_1 = \beta_2 = 0$. As far as the covariates are concerned, we choose two time-dependent covariates as given by

$$x_{it1} = \begin{cases} 0.0 & \text{for } i = 1,\ldots,K/2;\ t = 1,2 \\ 1.0 & \text{for } i = 1,\ldots,K/2;\ t = 3,T \\ 1.0 & \text{for } i = K/2+1,\ldots,K;\ t = 1,2 \\ 1.5 & \text{for } i = K/2+1,\ldots,K;\ t = 3,T, \end{cases}$$

and

$$x_{it2} = \begin{cases} 0.05+0.10(t-1) & \text{for } i = 1,\ldots,K/4;\ t = 1,\ldots,T \\ \frac{t}{4} & \text{for } i = K/4+1,\ldots,K/2;\ t = 1,\ldots,T \\ 0.0 & \text{for } i = K/2+1,\ldots,3K/4;\ t = 1,2, \\ 1.0 & \text{for } i = K/2+1,\ldots,3K/4;\ t = 3,T \\ -1.0 & \text{for } i = 3K/4+1,\ldots,K;\ t = 1,2, \\ 1.0 & \text{for } i = 3K/4+1,\ldots,K;\ t = 3,T. \end{cases}$$

Now by using these covariates and the generated data, we compute the estimates of β by (8.29), (8.30), and (8.31) under each of 1000 simulations. The simulated mean (SM), the simulated standard error (SSE), and the simulated mean squared error (SMSE) based on 1000 estimates for all three approaches are reported in Table 8.1.

The results of Table 8.1 show that for all selected values of σ_γ^2 and ρ, the GQL and CML approaches produce unbiased estimates with smaller standard errors (and hence smaller MSEs) as compared to the IVBGMM approach. Between the CML and GQL approaches, the CML approach produces unbiased estimates with larger SSEs than the GQL approach. Thus, the GQL approach clearly performs the best among the three approaches. Moreover, the GQL approach is also developed to estimate other parameters when needed. The IVBGMM approach at times appears to produce highly biased estimates, also with larger SSEs than the other two approaches. For example, when $\sigma_\gamma^2 = 1.0$, and $\rho = 0.5$, the GQL and CML approach

Table 8.1 Simulated mean, simulated standard error, and simulated mean squared error of the GQL, CML, and IVBGMM estimates for the regression parameters ($\beta = [\beta_1, \beta_2]'$) of AR(1) type Poisson mixed model (8.1) with known random effects variance ($\sigma_\gamma^2 = 0.5,\ 1.0$) and longitudinal correlation index parameter ($\rho = 0.50$); $K = 100$; $T = 4$; $\beta_1 = \beta_2 = 0.0$; 1000 simulations.

Variance Component (σ_γ^2)	Correlation Index Parameter (ρ)	Estimation Method	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$
0.50	0.50	GQL	SM	0.028	−0.013
			SSE	0.060	0.103
			SMSE	0.004	0.011
		CML	SM	0.014	0.036
			SSE	0.123	0.140
			SMSE	0.015	0.021
		IVBGMM	SM	0.133	−0.502
			SSE	1.069	1.318
			SMSE	1.160	1.989
1.00	0.50	GQL	SM	0.026	0.012
			SSE	0.066	0.127
			SMSE	0.005	0.016
		CML	SM	0.021	0.032
			SSE	0.109	0.138
			SMSE	0.012	0.020
		IVBGMM	SM	0.089	−0.481
			SSE	1.163	1.434
			SMSE	1.360	2.290

estimate $\beta_1 = 0$ as 0.026 and 0.021, respectively, whereas the IVBGMM approach yielded 0.089 as the estimate which is highly biased. The IVBGMM approach produced a much more biased estimate for β_2 as compared to the CML and GQL approaches. This is one of the reasons why we have computed the MSEs for comparison. For the same selected values of the parameters, that is, when $\sigma_\gamma^2 = 1.0$, and $\rho = 0.5$, the GQL, CML, and IVBGMM approaches produced the estimates of β_1 with MSEs 0.005, 0.012, and 1.360, respectively, and the estimates for β_2 with MSEs 0.016, 0.020, and 2.290, respectively. Thus, the IVBGMM approach cannot be trusted as it produces biased estimates with large standard errors, yielding large MSEs. When the SSEs or SMSEs for the GQL and CML are compared, the GQL approach appears to be the same or more efficient than the CML approach. For example, for the aforementioned estimation of β_1, the GQL approach is $0.012/0.005 = 2.4$ times more efficient (in the sense of MSE) than the CML approach. In summary, the GQL approach performs the best followed by the CML approach in estimating both β_1 and β_2.

8.2.2 Estimation of the Random Effects Variance σ_γ^2:

8.2.2.1 GMM Estimation for σ_γ^2

Note that because in the mixed model setup, that is, when $\rho = 0$ under the present longitudinal mixed model setup, conditional on γ_i, one can show that the count responses along with their squared and pairwise products are sufficient to estimate β and σ_γ^2 [Jiang (1998); Sutradhar (2004)], we now exploit such information to estimate the σ_γ^2 parameter. Note, however, that using the second-order information in the GMM approach, will require the fourth-order moments for the responses. The GQL estimation approach for σ_γ^2 discussed in the next section also requires these fourth-order moments. We now provide below the formulas for these fourth-order moments of the responses for the case when $\rho = 0$ [Sutradhar and Bari (2007); see also Section 4.2.6.2.] before we show the construction of the GMM estimating equation for σ_γ^2.

$$\begin{aligned}(a)\ E(Y_{it}^4|\rho=0) &= E_{\gamma_i}[E(Y_{it}^4|\gamma_i)]\\ &= E_{\gamma_i}[\mu_{it}^* + 7\mu_{it}^{*2} + 6\mu_{it}^{*3} + \mu_{it}^{*4}]\\ &= \mu_{it}\left[1 + 7\mu_{it}\exp(\sigma_\gamma^2) + 6\mu_{it}^2\exp(3\sigma_\gamma^2)\right.\\ &\quad \left.+\ \mu_{it}^3\exp(6\sigma_\gamma^2)\right]. \end{aligned} \tag{8.32}$$

$$\begin{aligned}(b)\ E(Y_{iu}^2Y_{it}^2|\rho=0) &= E_{\gamma_i}[E(Y_{iu}^2|\gamma_i)E(Y_{it}^2|\gamma_i)]\\ &= E_{\gamma_i}[\{\mu_{iu}^* + \mu_{iu}^{*\,2}\}\{\mu_{it}^* + \mu_{it}^{*2}\}]\\ &= \mu_{iu}\mu_{it}\exp(\sigma_\gamma^2)\left[1 + \{\mu_{iu} + \mu_{it}\}\exp(2\sigma_\gamma^2)\right.\\ &\quad \left.+\ \mu_{iu}\mu_{it}\exp(5\sigma_\gamma^2)\right]. \end{aligned} \tag{8.33}$$

$$\begin{aligned}(c)\ E(Y_{iu}^3Y_{it}|\rho=0) &= E_{\gamma_i}[E(Y_{iu}^3|\gamma_i)E(Y_{it}|\gamma_i)]\\ &= E_{\gamma_i}[\{\mu_{iu}^* + 3\mu_{iu}^{*\,2} + \mu_{iu}^{*\,3}\}\{\mu_{it}^*\}]\\ &= \mu_{iu}\mu_{it}\exp(\sigma_\gamma^2)\left[1 + 3\mu_{iu}\exp(2\sigma_\gamma^2)\right.\\ &\quad \left.+\ \mu_{iu}^2\exp(5\sigma_\gamma^2)\right]. \end{aligned} \tag{8.34}$$

$$(d)\ E(Y_{iu}^2Y_{iv}Y_{it}|\rho=0) = E_{\gamma_i}[E(Y_{iu}^2|\gamma_i)E(Y_{iv}|\gamma_i)E(Y_{it}|\gamma_i)]$$

$$= E_{\gamma_i}[\{\mu^*_{iu} + \mu^{*\,2}_{iu}\}\mu^*_{iv}\mu^*_{it}]$$

$$= \mu_{iu}\mu_{iv}\mu_{it}\exp(3\sigma^2_\gamma)[1 + \mu_{iu}\exp(3\sigma^2_\gamma)]. \quad (8.35)$$

$$(e)\ E(Y_{iu}Y_{iv}Y_{i\ell}Y_{it}|\rho = 0) = E_{\gamma_i}[E(Y_{iu}|\gamma_i)E(Y_{iv}|\gamma_i)E(Y_{i\ell}|\gamma_i)E(Y_{it}|\gamma_i)]$$

$$= E_{\gamma_i}[\mu^*_{iu}\mu^*_{iv}\mu^*_{i\ell}\mu^*_{it}]$$

$$= \mu_{iu}\mu_{iv}\mu_{i\ell}\mu_{it}\exp(6\sigma^2_\gamma). \quad (8.36)$$

In the fashion similar to that of the GMM estimation of β by solving (8.14), we obtain the GMM estimate for σ^2_γ by solving the GMM estimating equation

$$\frac{\partial \psi_2}{\partial \sigma^2_\gamma} C_2 \psi_2 = 0, \quad (8.37)$$

where $\psi_2 = K^{-1}\sum_{i=1}^K \psi_{2i}$ with $\psi_{2i} = \sum_{u\le t}^T [y_{iu}y_{it} - \lambda_{iut}]$, λ_{iut} being the expectation of the second-order responses $Y_{iu}Y_{it}$ for $u \le t$. Note that by (8.6) and (8.8), we obtain

$$\lambda_{iut} = E[Y_{iu}Y_{it}] = \begin{cases} \mu_{it} + [\exp(\sigma^2_\gamma)]\mu^2_{it}, & \text{for } u = t \\ \rho^{t-u}\mu_{iu} + [\exp(\sigma^2_\gamma)]\mu_{iu}\mu_{it} & \text{for } u < t, \end{cases} \quad (8.38)$$

so that

$$\frac{\partial \psi_2}{\partial \sigma^2_\gamma} = K^{-1}\sum_{i=1}^K \left[\sum_{t=1}^T [\mu_{it}/2 + 2\exp(\sigma^2_\gamma)\mu^2_{it}] \right.$$

$$\left. + \sum_{u<t}^T [\rho^{t-u}\mu_{iu}/2 + 2\exp(\sigma^2_\gamma)\mu_{iu}\mu_{it}] \right].$$

Note that C_2 in (8.37) is not easy to compute for general ρ. We rather compute this matrix by using a 'working' longitudinal independence assumption; that is, $\rho = 0$. The formula for C_2 is then given by

$$C_2^{-1} = \text{var}\left(K^{-1}\sum_{i=1}^K \psi_{2i} \right)$$

$$= K^{-2}\left[\sum_{i=1}^K \sum_{u\le \ell}^T \sum_{m\le t}^T \tilde{\phi}_{iu\ell mt}(\beta,\ \sigma^2_\gamma,\ \rho = 0) \right], \quad (8.39)$$

where for $u \le \ell$ and $m \le t$, the formula for

$$\tilde{\phi}_{iu\ell mt}(\beta,\ \sigma^2_\gamma,\ \rho = 0) = \text{cov}[Y_{iu}Y_{i\ell}, Y_{im}Y_{it}|\rho = 0]$$

$$= E[Y_{iu}Y_{i\ell}Y_{im}Y_{it}|\rho = 0] - E[Y_{iu}Y_{i\ell}|\rho = 0]E[Y_{im}Y_{it}|\rho = 0]$$

$$= E[Y_{iu}Y_{i\ell}Y_{im}Y_{it}|\rho = 0] - \lambda_{iu\ell}|_{\rho=0}\lambda_{imt}|_{\rho=0}$$

can be computed by using the fourth-order moments formula given from (8.32) to (8.36), and the second-order expectation given by (8.38). For example, when $u = \ell$ and $m = t$ but $u \neq m$, the value for the fourth-order moment $E[Y_{iu}^2Y_{im}^2|\rho = 0]$ would be computed by (8.33).

8.2.2.2 GQL Estimation for σ_γ^2:

For the estimation of σ_γ^2, the GQL approach exploits the squared and the pairwise product of the observations in a different manner from the MM approach (8.37). Let

$$u_i = (y_{i1}^2, \ldots, y_{iT}^2, y_{i1}y_{i2}, \ldots, y_{it}y_{i,t+1}, \ldots, y_{i,T-1}y_{iT})',$$

and write its expectation as

$$\lambda_i(\beta, \sigma_\gamma^2, \rho) = (\lambda_{i11}, \ldots, \lambda_{itt}, \ldots, \lambda_{iTT}, \lambda_{i12}, \ldots, \lambda_{iut}, \ldots, \lambda_{i,T-1,T})',$$

where by (8.6), one obtains

$$\lambda_{itt} \equiv \lambda_{itt}(\beta, \sigma_\gamma^2) = E(Y_{it}^2) = \mu_{it} + \mu_{it}^2 \exp(\sigma_\gamma^2), \tag{8.40}$$

for all $t = 1, \ldots, T$, and by (8.8)

$$\lambda_{iut} \equiv \lambda_{iut}(\beta, \sigma_\gamma^2, \rho) = E(Y_{iu}Y_{it}) = \rho^{t-u}\mu_{iu} + \mu_{iu}\mu_{it}\exp(\sigma_\gamma^2), \tag{8.41}$$

for all $u < t$. By using the QL principle similar to that of (8.18), one may now write the GQL estimating equation for σ_γ^2 as

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta, \sigma_\gamma^2, \rho)}{\partial \sigma_\gamma^2} \Omega_i^{-1}(\beta, \sigma_\gamma^2, \rho)[u_i - \lambda_i(\beta, \sigma_\gamma^2, \rho)] = 0, \tag{8.42}$$

[Sutradhar and Jowaheer (2003)] where Ω_i is the covariance matrix of u_i. Note that it is, however, extremely cumbersome to compute Ω_i in general under the auto-regression model (8.1). As a remedy, one may use a 'working' covariance matrix of u_i such as $\Omega_{iw}(\beta, \sigma_\gamma^2, \rho = 0)$ under the 'working' assumption of conditional longitudinal independence; that is,

$$\text{corr}(Y_{iu}, Y_{it}|\gamma_i) = 0, \tag{8.43}$$

whereas the true conditional correlation is assumed to be given by (8.3). Thus we propose to solve the 'working' GQL estimating equation

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta, \sigma_\gamma^2, \rho)}{\partial \sigma_\gamma^2} \Omega_{iw}^{-1}(\beta, \sigma_\gamma^2, \rho = 0)[u_i - \lambda_i(\beta, \sigma_\gamma^2, \rho)] = 0, \tag{8.44}$$

for the estimation of σ_γ^2. Note that the GQL estimate of σ_γ^2 obtained from (8.44) may not be highly efficient because of the use of a 'working' weight matrix. This estimator is, however, consistent as (8.44) is an unbiased estimating equation.

Computation of $\Omega_{iw}(\beta, \sigma_\gamma^2, \rho = 0)$

For the computation of the elements of the 'working' higher-order moments matrix $\Omega_{iw}(\beta, \sigma_\gamma^2, \rho = 0)$ in (8.44), we need the formulas for the fourth-order moments of the responses under the assumption that $\rho = 0$. These fourth-order moments are given in the equations from (8.32) to (8.36), and they may now be used to compute the elements of the $\Omega_{iw}(\beta, \sigma_\gamma^2, \rho = 0)$ matrix. For example, we use the formula for $E[Y_{iu}^2 Y_{iv} Y_{it} | \rho = 0]$ from (8.35) and compute

$$\begin{aligned} \text{cov}[(Y_{iu}^2, Y_{iv}Y_{it})|\rho = 0] &= E[Y_{iu}^2 Y_{iv} Y_{it} | \rho = 0] - E[Y_{iu}^2] E[Y_{iv} Y_{it} | \rho = 0] \\ &= \mu_{iu}\mu_{iv}\mu_{it} \exp(3\sigma_\gamma^2)[1 + \mu_{iu} \exp(3\sigma_\gamma^2)] \\ &\quad -[\mu_{iu} + \mu_{iu}^2 \exp(\sigma_\gamma^2)] \left\{ \mu_{iv}\mu_{it} \exp(\sigma_\gamma^2) \right\}. \end{aligned} \tag{8.45}$$

8.2.2.3 Asymptotic Efficiency Comparison : GMM versus GQL

The means and the variances of the count responses are functions of both β and σ_γ^2, thus we are mainly interested in examining the relative efficiency of the GMM and GQL approaches in estimating these parameters. Below we provide the covariance matrix of the estimator of β and the variance of the estimator of σ_γ^2, under the GMM and GQL approaches.

8.2.2.3.1 Asymptotic Variances of the GMM Estimators

The K individuals provide repeated count responses independently, thus as $K \to \infty$, it follows from (8.14) by using the central limit theorem that the GMM based estimator of β has the asymptotic covariance matrix given by

$$\text{cov}(\hat{\beta}_{GMM}) = \text{Lt}_{K\to\infty} \left[\frac{\partial \psi_1'}{\partial \beta} C_1 \frac{\partial \psi_1}{\partial \beta'} \right]^{-1}. \tag{8.46}$$

Similarly, by (8.37), one derives the asymptotic variance of the GMM estimator of σ_γ^2 as

$$\text{var}(\hat{\sigma}_{\gamma,GMM}^2) = \text{Lt}_{K\to\infty} C_2^{-1} \left[\frac{\partial \psi_2}{\partial \sigma_\gamma^2} \right]^{-2}. \tag{8.47}$$

8.2.2.3.2 Asymptotic Variances of the GQL Estimators

By similar calculations as in the GMM case, it follows from (8.18) that the asymptotic covariance matrix of the GQL estimator of β is given by

$$\mathrm{cov}(\hat{\beta}_{GQL}) = \mathrm{Lt}_{K\to\infty}\left[\sum_{i=1}^{K}\frac{\partial\mu_i'(\beta,\sigma_\gamma^2)}{\partial\beta}\Sigma_i^{-1}(\beta,\sigma_\gamma^2,\rho)\frac{\partial\mu_i(\beta,\sigma_\gamma^2)}{\partial\beta'}\right]^{-1}. \tag{8.48}$$

Next, it follows from (8.44) that the asymptotic variance of the GQL estimator of σ_γ^2 is given by

$$\begin{aligned}\mathrm{var}(\hat{\sigma}^2_{\gamma,GQL}) = \mathrm{Lt}_{K\to\infty}&\left[\sum_{i=1}^{K}\frac{\partial\lambda_i'(\beta,\sigma_\gamma^2,\rho)}{\partial\sigma_\gamma^2}\Omega_{iw}^{-1}(\beta,\sigma_\gamma^2,\rho=0)\frac{\partial\lambda_i(\beta,\sigma_\gamma^2,\rho)}{\partial\sigma_\gamma^2}\right]^{-2}\\ &\times\left[\sum_{i=1}^{K}\frac{\partial\lambda_i'(\beta,\sigma_\gamma^2,\rho)}{\partial\sigma_\gamma^2}\Omega_{iw}^{-1}(\beta,\sigma_\gamma^2,\rho=0)\Omega_i(\beta,\sigma_\gamma^2,\rho)\right.\\ &\times\left.\Omega_{iw}^{-1}(\beta,\sigma_\gamma^2,\rho=0)\frac{\partial\lambda_i(\beta,\sigma_\gamma^2,\rho)}{\partial\sigma_\gamma^2}\right].\end{aligned} \tag{8.49}$$

8.2.2.3.3 Asymptotic Efficiency Computation

One may now compute the asymptotic efficiency of the GQL estimators for the components of β as compared to those of the GMM estimators by comparing the respective diagonal elements of the covariance matrices given in (8.46) and (8.48). Similarly, the asymptotic efficiency of the GQL estimator of σ_γ^2 to the GMM estimator is found by comparing the variances in (8.47) and (8.49).

We now illustrate the relative efficiency of the GQL and GMM estimators through a numerical example. For the purpose, we consider $K = 500,\ p = 2,\ T = 4$, and use a covariate matrix with the first covariate as

$$x_{it1} = \begin{cases} 0 \;\; \text{for } i = 1,\ldots,K/2;\ t = 1,2 \\ 1 \;\; \text{for } i = 1,\ldots,K/2;\ t = 3,4 \\ 1 \;\; \text{for } i = K/2+1,\ldots,K;\ t = 1,\ldots,4, \end{cases}.$$

whereas the second covariate is chosen to be

$$x_{it2} = \begin{cases} 1 \quad\;\; \text{for } i = 1,\ldots,K/2;\ t = 1,2 \\ 1.5 \;\; \text{for } i = 1,\ldots,K/2;\ t = 3,4 \\ 0 \quad\;\; \text{for } i = K/2+1,\ldots,K;\ t = 1,2 \\ 1 \quad\;\; \text{for} i = K/2+1,\ldots,K;\ t = 3,4. \end{cases}.$$

Furthermore, for true parameter values, we consider $\beta_1 = \beta_2 = 1.0$, $\rho = 0.8$, and $\sigma_\gamma^2 = 0.5$, and 1.5. The asymptotic variances computed from (8.46) through (8.49) are shown in Table 8.2 for the above selection of parameter values.

Table 8.2 Comparison of asymptotic variances of the GQL and GMM estimators for the estimation of the regression parameters (β_1 and β_2), and the variance component (σ_γ^2), of a longitudinal mixed model for count panel data, with $T = 4$, $\rho = 0.8$, and $K = 500$.

		Asymptotic Variances	
Method	Quantity	$\sigma_\gamma^2 = 0.5$	1.5
GQL	$\text{Var}(\hat{\beta}_1)$	9.68×10.0^{-4}	6.62×10.0^{-4}
	$\text{Var}(\hat{\beta}_2)$	6.74×10.0^{-4}	4.86×10.0^{-4}
	$\text{Var}(\hat{\sigma}_\gamma^2)$	7.38×10.0^{-5}	1.08×10.0^{-5}
GMM	$\text{Var}(\hat{\beta}_1)$	4.26×10.0^{-3}	1.93×10.0^{-2}
	$\text{Var}(\hat{\beta}_2)$	3.16×10.0^{-3}	1.49×10.0^{-2}
	$\text{Var}(\hat{\sigma}_\gamma^2)$	3.74×10.0^{-3}	0.230

The results of Table 8.2 show that the variances of the estimators for all three main parameters β_1, β_2, and σ_γ^2, under the GQL approach are uniformly much smaller than the corresponding variances under the GMM approach, justifying that the GQL approach produces much more efficient estimates than the GMM approach for all main parameters of the model. For example, when $\sigma_\gamma^2 = 1.5$, the GQL estimates of β_1 and β_2 are, respectively,

$$\frac{1.93 \times 10.0^{-2}}{6.62 \times 10.0^{-4}} = 29.15 \text{ and } \frac{1.49 \times 10.0^{-2}}{4.86 \times 10.0^{-4}} = 30.66$$

times more efficient than the corresponding GMM estimates. For the estimation of σ_γ^2, the GQL approach appears to perform extra-ordinarily better than the GMM approach. For example, for the same set of parameters, (i.e., when $\rho = 0.8$ and $\sigma_\gamma^2 = 1.5$), the GQL estimate of σ_γ^2 is

$$\frac{0.230}{1.08 \times 10.0^{-5}} = 21296.30$$

times more efficient. In summary, the GQL approach performs much better than the GMM approach in estimating all main parameters, its performance being extraordinarily better in estimating the variance component σ_γ^2.

8.2.3 Estimation of the Longitudinal Correlation Parameter ρ

8.2.3.1 GMM Estimation for ρ

In the fashion similar to that of the GMM estimation of σ_γ^2 by solving (8.37), we obtain the GMM estimate for ρ by solving the estimating equation

$$\frac{\partial \psi_3'}{\partial \rho} C_3 \psi_3 = 0, \tag{8.50}$$

where $\psi_3 = K^{-1} \sum_{i=1}^{K} \psi_{3i}$ with $\psi_{3i} = \sum_{t=2}^{T} [y_{it} y_{i,t-1} - \lambda_{it,t-1}]$, where by (8.38)

$$\lambda_{it,t-1} = E[Y_{it} Y_{i,t-1}] = \rho \mu_{i,t-1} + [\exp(\sigma_\gamma^2)] \mu_{i,t-1} \mu_{it}, \tag{8.51}$$

so that

$$\frac{\partial \psi_3}{\partial \rho} = K^{-1} \sum_{i=1}^{K} \sum_{t=2}^{T} [\mu_{i,t-1}]. \tag{8.52}$$

Similar to the computation for C_2 in (8.37), we compute C_3 in (8.50) under the 'working' longitudinal independence assumption ($\rho = 0$). To be specific,

$$\begin{aligned} C_3^{-1} &= \text{var}\left(K^{-1} \sum_{i=1}^{K} \psi_{3i} \right) \\ &= K^{-2} \left[\sum_{i=1}^{K} \sum_{u=2}^{T} \sum_{t=2}^{T} \phi_{iu,u-1,t,t-1}(\beta,\, \sigma_\gamma^2,\, \rho = 0) \right. \\ &\quad \left. - \left(\sum_{i=1}^{K} \sum_{t=2}^{T} \lambda_{it,t-1}(\beta,\, \sigma_\gamma^2,\, \rho = 0) \right)^2 \right], \end{aligned} \tag{8.53}$$

where $\lambda_{it,t-1}$ is given by (8.51), and the formula for $\phi_{iu,u-1,t,t-1}(\beta,\, \sigma_\gamma^2,\, \rho = 0)$ can be computed by using the fourth-order moments formula given from (8.32) to (8.36). Note that this moment estimate can be inefficient because the weight matrix C_3 is computed by pretending that $\rho = 0$, when ρ itself is the parameter of interest.

8.2.3.2 ρ Estimation Under the GQL Approach

Note that the regression effect β and the variance of the random effects have been estimated by using the GQL and the 'working' GQL estimating equations (8.18) and (8.44), respectively, for a known value of ρ. But in practice ρ is rarely known. For given β and σ_γ^2, the correlation or probability parameter (ρ) may be consistently estimated by solving a suitable moment estimating equation that may be developed by equating the population covariance of the data given in (8.8) with its sample counterpart. Note that as ρ is a correlation parameter under the autoregressive order 1 setup, similar to the Gaussian setup, it would be sufficient to exploit the lag 1 autocovariance only to estimate this parameter. More specifically, as by (8.6)

$$E(Y_{it} - \mu_{it})^2 = \sigma_{itt} = \mu_{it} + [\exp(\sigma_\gamma^2) - 1]\mu_{it}^2,$$

and by (8.8)

$$E(Y_{it}-\mu_{it})(Y_{i,t+1}-\mu_{i,t+1}) = \rho\mu_{it} + \{\exp(\sigma_\gamma^2)-1\}\mu_{it}\mu_{i,t+1},$$

ρ may be estimated consistently by

$$\hat{\rho} = \frac{a_1 - b_1}{g_1}, \tag{8.54}$$

where a_1 is the observed lag 1 correlation defined as

$$a_1 = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-1} y_{it}^* y_{i(t+1)}^* / K(T-1)}{\sum_{i=1}^{K}\sum_{t=1}^{T} y_{it}^{*2} / KT},$$

with $y_{it}^* = (y_{it}-\mu_{it})/(\sigma_{itt})^{1/2}$, where $\sigma_{itt} = \mu_{it} + (\exp(\sigma_\gamma^2)-1)\mu_{it}^2$. In (8.54),

$$g_1 = \sum_{i=1}^{K}\sum_{t=1}^{T-1} \mu_{it}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}/K(T-1),$$

and

$$b_1 = (\exp(\sigma_\gamma^2)-1)\sum_{i=1}^{K}\sum_{t=1}^{T-1} m_{it}m_{i,t+1}/K(T-1),$$

with $m_{it} = \mu_{it}/(\sigma_{itt})^{1/2}$.

Note that at every stage of iterations the estimate of ρ from (8.54) must satisfy the range restriction (8.4). In case the estimate falls outside the boundary of this range in a given iteration, we use the upper or lower limit as appropriate as the estimate under that iteration.

8.2.4 A Simulation Study

To examine the finite sample performance of the GQL approach for the estimation of all three parameters, namely β, σ_γ^2, and ρ, Sutradhar and Bari (2007) conducted an extensive simulation study for both balanced and unbalanced designs. The data were generated following the model (8.1) for $K = 100$ individuals based on $p = 2$ covariates with their effects $\beta_1 = 0.0$, $\beta_2 = 0.0$; and for various values of $\sigma_\gamma^2 = 0.25,\ 0.5,\ 0.75,$ and 1.0; and $\rho = 0.25,\ 0.5,$ and 0.75. We consider here $\sigma_\gamma^2 = 0.5,$ and 1.0; and $\rho = 0.5,$ and 0.75.

A Balanced Design (D_1)

As far as the covariates are concerned, they are available from all $K = 100$ individuals for all time points $t = 1,\ldots,4$. The covariates are chosen as

$$x_{it1} = \begin{cases} 1/2 & \text{for } i = 1,\ldots,K/4;\ t = 1,2 \\ 0 & \text{for } i = 1,\ldots,K/4;\ t = 3,4 \\ -1/2 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 1 \\ 0 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 2,3 \\ 1/2 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 4 \\ t/8 & \text{for } i = 3K/4+1,\ldots,K;\ t = 1,\ldots,4, \end{cases},$$

$$Pr(x_{it2} = 1) = \begin{cases} 0.3 & \text{for } t = 1 \\ 0.5 & \text{for } t = 2,3 \\ 0.8 & \text{for } t = 4. \end{cases}.$$

We now generate the data following the model (8.1) based on the aforementioned covariates and parameter values. Consider 1000 simulations. In each simulation, for the full model (8.1), we obtain the GQL estimates for $\beta = (\beta_1, \beta_2)'$ and σ_γ^2 by using the estimating equations (8.18) and (8.44), respectively. The correlation parameter (ρ) is estimated by (8.54). The simulated mean, simulated standard error, and simulated relative bias (SRB) computed by

$$\text{SRB} = \frac{\text{Absolute Bias}}{\text{SSE}} \times 100$$

for the full model are reported in Table 8.3.

Table 8.3 Simulated mean, simulated standard error, and simulated relative bias of the GQL estimates for parameters of the nonstationary longitudinal mixed model with the balanced design (D_1), for selected values of σ_γ^2 and ρ; $K = 100$; $T = 4$; $\beta_1 = \beta_2 = 0.0$; 1000 simulations.

Variance Component (σ_γ^2)	Correlation Parameter(ρ)	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\gamma^2$	$\hat{\rho}$
0.50	0.50	SM	0.028	−0.014	0.424	0.522
		SSE	0.148	0.089	0.186	0.094
		SRB	19	16	41	23
	0.75	SM	0.012	−0.014	0.430	0.754
		SSE	0.117	0.088	0.202	0.063
		SRB	10	16	35	6
1.0	0.50	SM	0.017	−0.008	0.833	0.500
		SSE	0.162	0.089	0.244	0.135
		SRB	11	9	68	0
	0.75	SM	−0.011	0.000	0.823	0.746
		SSE	0.124	0.091	0.222	0.086
		SRB	9	0	80	1

An Unbalanced Design (D_2)

Note that even though the GQL estimating equations for β (8.18) and σ^2_γ (8.44), and the moment estimation of ρ by (8.54) were developed for balanced longitudinal data, they can easily be modified for unbalanced data. More specifically, the estimating equations for β (8.18) and σ^2_γ (8.44) may be adjusted for the unbalanced case by using T_i for T for the ith $(i = 1,\ldots,K)$ individual. The moment estimating formula (8.54) for ρ requires minor adjustment for all three quantities a_1, b_1, and g_1. Thus, under the present unbalanced case, the formula for a_1, b_1, and g_1 may be written as

$$\begin{aligned}
a_1 &= \frac{\sum_{i=1}^{K}\sum_{t=1}^{T_i-1} y^*_{it}y^*_{i(t+1)} / \sum_{i=1}^{K}(T_i-1)}{\sum_{i=1}^{K}\sum_{t=1}^{T_i} y^{*2}_{it} / \sum_{i=1}^{K} T_i} \\
b_1 &= (\exp(\sigma^2_\gamma)-1)\sum_{i=1}^{K}\sum_{t=1}^{T_i-1} m_{it}m_{i,t+1} / \sum_{i=1}^{K}(T_i-1) \\
g_1 &= \sum_{i=1}^{K}\sum_{t=1}^{T_i-1} \mu_{it}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-\frac{1}{2}} / \sum_{i=1}^{K}(T_i-1),
\end{aligned} \tag{8.55}$$

with $m_{it} = \mu_{it}/(\sigma_{itt})^{1/2}$. Also note that the GQL estimation technique is not restricted to any particular values of the regression parameters. Thus, in this simulation study for the unbalanced data, we choose nonzero regression effects, namely $\beta_1 = 1.0$ and $\beta_2 = 0.5$. The values of σ^2 and ρ remain the same, for example, as in the last simulation study conducted for the balanced data. As far as the unbalanced design for the covariates is concerned, we choose them as

$$x_{it1} = \begin{cases} -1.0 & \text{for } i = 1,\ldots,K/2;\ t = 1,2 \\ 1.0 & \text{for } i = 1,\ldots,K/2;\ t = 3,4 \\ -1/2 & \text{for } i = K/2+1,\ldots,3K/4;\ t = 1 \\ 0 & \text{for } i = K/2+1,\ldots,3K/4;\ t = 2,3 \\ 1/2 & \text{for } i = K/2+1,\ldots,3K/4;\ t = 4 \\ t/6 & \text{for } i = 3K/4+1,\ldots,K;\ t = 1,\ldots,3, \end{cases}.$$

and

$$Pr(x_{it2}=1)=\begin{cases}0.3 & \text{for } i=1,\ldots,3K/4;\ t=1\\ 0.5 & \text{for } i=1,\ldots,3K/4;\ t=2,3\\ 0.8 & \text{for } i=1,\ldots,3K/4;\ t=4,\\ 0.3 & \text{for } i=3K/4+1,\ldots,K;\ t=1,\\ 0.5 & \text{for } i=3K/4+1,\ldots,K;\ t=2,\\ 0.8 & \text{for } i=3K/4+1,\ldots,K;\ t=3,\end{cases}.$$

The simulation results for the present unbalanced design case are reported in Table 8.4. Note that in this simulation study, we have used $\sum_{i=1}^{K} T_i = 375$, whereas in the balanced design case, we considered $KT = 400$.

Table 8.4 Unbalanced design (D_2) based simulated mean, simulated standard error, estimated standard error, and simulated relative bias of the GQL estimates for parameters of the nonstationary longitudinal mixed model with two covariates for selected values of σ_γ^2 and ρ; $K = 100$; $\beta_1 = 1.0$ and $\beta_2 = 0.5$; 1000 simulations.

Variance Component (σ_γ^2)	Correlation Parameter(ρ)	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\gamma^2$	$\hat{\rho}$
0.50	0.50	SM	0.971	0.510	0.478	0.458
		SSE	0.146	0.086	0.192	0.098
		ESE	0.178	0.109	0.279	–
		SRB	20	12	11	43
	0.75	SM	0.960	0.516	0.498	0.695
		SSE	0.134	0.091	0.211	0.102
		ESE	0.184	0.112	0.310	–
		SRB	30	17	1	54
1.0	0.50	SM	0.955	0.493	0.810	0.469
		SSE	0.150	0.093	0.206	0.091
		ESE	0.208	0.131	0.362	–
		SRB	30	8	92	35
	0.75	SM	0.949	0.501	0.811	0.676
		SSE	0.132	0.095	0.199	0.114
		ESE	0.184	0.130	0.316	–
		SRB	38	1	95	65

With regard to the estimation of the regression effect, both Tables 8.3 and 8.4 show that the GQL estimating equation (8.18) appears to perform well. To estimate the overdispersion parameter (σ_γ^2), the GQL estimating equation (8.44) appears to underestimate this parameter, the estimates being better for small values of σ_γ^2. The moment estimation for the longitudinal correlation parameter ρ appears to work extremely well in the balanced data case (8.54), whereas in the unbalanced case, the moment formula (8.55) appears to underestimate this correlation parameter ρ. For example, for $\sigma_\gamma^2 = 0.5$ and $\rho = 0.75$, the balanced design based results from

Table 8.3 show that the estimates of $\beta_1 = 0.0$, $\beta_2 = 0.0$, σ^2_γ, and ρ are found to be 0.012, −0.014, 0.430, and 0.754, respectively with relative bias 10, 16, 35, and 6. Similarly, for the same values of σ^2_γ and ρ, the unbalanced design based results from Table 8.4 show that the estimates of $\beta_1 = 1.0$, $\beta_2 = 0.5$, σ^2_γ, and ρ are found to be 0.960, 0.516, 0.498, and 0.695, respectively, with relative bias 30, 17, 1, and 54.

Note that because the exact computation of the fourth-order weight matrix Ω_i in (8.42) is either impossible or extremely difficult, we have approximated it by a 'working' conditional independence ($\rho = 0$) assumption based weight matrix, for the estimation of σ^2_γ by (8.44). This approximation yielded slightly biased estimates both in Tables 8.3 and 8.4, especially when the true value of σ^2_γ was large. Some authors have used a different approximation such as 'working' multivariate normality based approximation to the weight matrix, where, for the purpose of computing the weight matrix only, it is pretended that the repeated responses of an individual follow the multivariate normal distribution with correct count mean vector and covariance matrix. See, for example, Jowaheer and Sutradhar (2002) [see also Prentice and Zhao (1991) in the context of fixed longitudinal models] for such an approximation under a longitudinal model for negative binomial count data. This normality based approximation for the improvement of σ^2_γ estimation is discussed further in Section 8.3 in the context of a conditional serially correlated model with a deflated marginal mean as compared to the marginal mean (8.5) under the model (8.1).

We now turn back to examine how different simpler 'working' approximations can negatively affect the estimation of nonregression parameters. More specifically, in Section 8.2.4.1, we check through a simulation study, the model misspecification effect of completely ignoring ρ on the estimation of β and σ^2_γ. Similarly in Section 8.2.4.2, we conduct another simulation study to examine the model misspecification effect of completely ignoring σ^2_γ on the estimation of β and ρ.

8.2.4.1 Estimation Under the 'Working' Conditional Independence ($\rho = 0$) Model

The purpose of this section is to examine the effect of using the assumption $\rho = 0$ in estimating both β and σ^2_γ parameters. Note that this is different from the estimation of σ^2_γ by using 'working' independence based fourth-order matrix as a weight in the GQL estimating equation (8.44). Here the data are generated under the full model (8.1), but the parameters β and σ^2_γ are estimated by writing the GQL estimating equations for β and σ^2_γ under the conditional independence ($\rho = 0$) assumption.

As far as the estimating equations for β and σ^2_γ under the condition $\rho = 0$, are concerned, the 'working' independence based GQL estimating equation for β, following (8.18), may be written as

$$\sum_{i=1}^{K} \frac{\partial \mu_i'(\beta, \sigma^2_\gamma)}{\partial \beta} \Sigma_i^{-1}(\beta, \sigma^2_\gamma, \rho = 0)[y_i - \mu_i(\beta, \sigma^2_\gamma)] = 0. \tag{8.56}$$

Similarly, the 'working' independence based GQL estimating equation for σ^2_γ, following (8.44), may be written as

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta, \sigma^2_\gamma, \rho = 0)}{\partial \sigma^2_\gamma} \Omega_{iw}^{-1}(\beta, \sigma^2_\gamma, \rho = 0)[u_i - \lambda_i(\beta, \sigma^2_\gamma, \rho = 0)] = 0. \qquad (8.57)$$

Note that the $\Sigma_i(\beta, \sigma^2_\gamma, \rho = 0)$ in (8.56) may easily be computed by putting $\rho = 0$ in the formulas for the elements of the $\Sigma_i(\beta, \sigma^2_\gamma, \rho)$ matrix defined in (8.18). Similarly, the formulas for $\lambda_i(\beta, \sigma^2_\gamma, \rho = 0)$ in (8.44) may be computed by putting $\rho = 0$ in the formula for $\lambda_{iut}(\beta, \sigma^2_\gamma, \rho)$ defined in (8.41). Furthermore, the fourth-order moments based weight matrix $\Omega_{iw}(\beta, \sigma^2_\gamma, \rho = 0)$ in (8.57) is the same as that of (8.44), and hence no additional computation is required.

Table 8.5 [Estimation effects when ρ is ignored ($\rho = 0$)**]** Simulated mean, simulated standard error, and simulated relative bias of the GQL estimates for parameters of the pretended nonstationary mixed model with two covariates for selected values of σ^2_γ and ρ; $K = 100$; $T = 4$; $\beta_1 = \beta_2 = 0.0$; 1000 simulations.

Convergent Simulations	Variance Component (σ^2_γ)	Correlation Parameter(ρ)	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2_\gamma$
68	0.50	0.50	SM	0.042	−0.028	1.099
			SSE	0.138	0.095	0.261
			SRB	30	30	330
0		0.75	SM	–	–	–
			SSE	–	–	–
			SRB	–	–	–
41	1.0	0.50	SM	0.013	−0.012	1.879
			SSE	0.167	0.080	0.935
			SRB	8	15	94
5		0.75	SM	0.034	−0.063	3.601
			SSE	0.127	0.066	0.148
			SRB	27	96	1757

Now by generating the data under the full model (8.1), but by computing β and σ^2_γ from (8.56) and (8.57), respectively, one obtains the misspecified model based estimates. These estimates are reported in Table 8.5. Note that the results of the table indicate that the conditional independence ($\rho = 0$) assumption encountered serious convergence problems. To be specific, when the true ρ is large but it was considered to be 0 (i.e., $\rho = 0$), there are almost no convergent simulations. This shows that one may not be able to estimate β and σ^2_γ by ignoring the longitudinal correlation parameter. In simulations, particularly for small ρ, where the convergence problem was not so serious, the estimates of β_1 and β_2 are found to be satisfactory, whereas the estimates of σ^2_γ are found to be highly positively biased and hence not trustworthy.

8.2.4.2 Estimation Under the 'Working' Longitudinal Fixed ($\sigma_\gamma^2 = 0$) Model

In this section, we examine the effect of ignoring the random effects on the estimation of β and ρ. To be specific, even though the responses are generated in the presence of β, σ_γ^2, and ρ, we, however, estimate the β and ρ parameters under the assumption that $\sigma_\gamma^2 = 0$. In this case, the regression parameter β is estimated by solving the GQL estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \mu_i'(\beta, \sigma_\gamma^2 = 0)}{\partial \beta} \Sigma_i^{-1}(\beta, \sigma_\gamma^2 = 0, \rho)[y_i - \mu_i(\beta, \sigma_\gamma^2 = 0)] = 0, \tag{8.58}$$

where $\mu_i(\beta, \sigma_\gamma^2 = 0)$ and $\Sigma_i(\beta, \sigma_\gamma^2 = 0, \rho)$ are obtained by evaluating $\mu_i(\beta, \sigma_\gamma^2)$ and $\Sigma_i(\beta, \sigma_\gamma^2, \rho)$ with $\sigma_\gamma^2 = 0$.

For the estimation of ρ when $\sigma_\gamma^2 = 0$, we still can use the formula

$$\hat{\rho} = \frac{a_1 - b_1}{g_1},$$

following (8.54), but a_1, b_1, and g_1 have to be evaluated at $\sigma_\gamma^2 = 0$. For example,

$$a_1(\sigma_\gamma^2 = 0) = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-1} \tilde{y}_{it} \tilde{y}_{i(t+1)} / K(T-1)}{\sum_{i=1}^{K} \sum_{t=1}^{T} \tilde{y}_{it}^2 / KT},$$

where $\tilde{y}_{it} = (y_{it} - \mu_{it|\sigma_\gamma^2=0}) / \sigma_{itt|\sigma_\gamma^2=0}^{1/2}$. The simulation results are given in Table 8.6.

Table 8.6 [Estimation effects when σ_γ^2 is ignored ($\sigma_\gamma^2 = 0$)] Simulated mean, simulated standard error, and simulated relative bias of the GQL estimates for parameters of the pretended nonstationary longitudinal fixed model with two covariates for selected values of σ_γ^2 and ρ; $K = 100$; $T = 4$; $\beta_1 = \beta_2 = 0.0$; 1000 simulations.

Variance Component (σ_γ^2)	Correlation Parameter(ρ)	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
0.50	0.50	SM	0.040	−0.018	0.683
		SSE	0.145	0.089	0.058
		SRB	28	20	316
	0.75	SM	0.015	−0.017	0.836
		SSE	0.115	0.088	0.043
		SRB	13	19	200
1.0	0.50	SM	0.023	−0.012	0.775
		SSE	0.150	0.087	0.049
		SRB	15	14	561
	0.75	SM	−0.008	−0.007	0.884
		SSE	0.121	0.092	0.037
		SRB	7	8	362

The results of Table 8.6 show that the estimates of β_1 and β_2 are slightly biased, whereas the SRB for the estimate of ρ is much larger as compared to that of Table 8.3. For example, when the true value of σ_γ^2 is 1.0 but the estimates of $\beta_1 = 0.0$, $\beta_2 = 0.0$, and $\rho = 0.75$ are obtained under the fixed model assumption (i.e., $\sigma_\gamma^2 = 0$), the SRBs for β_1, β_2, and ρ are found to be 7, 8, and 362, respectively, whereas the corresponding SRBs are 9, 0, and 1, respectively, under the full model (Table 8.3). This shows significant detrimental effects on the estimation of ρ due to ignoring the σ_γ^2.

When the results of these two Tables 8.5 and 8.6 are summarized, it appears that the model misspecification has detrimental effects mainly on the estimation of nonregression parameters. More specifically, the estimation of σ_γ^2 (Table 8.5) by ignoring the longitudinal correlations appears to have serious nonconvergence problems.

8.2.5 An Illustration: Analyzing Health Care Utilization Count Data by Using Longitudinal Fixed and Mixed Models

In Chapter 6 (see Section 6.7), a health care utilization dataset was analyzed by fitting a fixed longitudinal model under the assumption that the four repeated count responses from each of 144 individuals follow a Poisson AR(1) type correlation model. Note that the health care utilization data given in Appendix 6A contains the record of the number of physician visits by 180 individuals. Among them, each of the 36 additional (to what was considered in Section 6.7) individuals had a record of repeated count responses for three years. Thus, the data are unbalanced, whereas in Section 6.7 we analyzed the larger balanced segment of the data. Sutradhar and Bari (2007) have reanalyzed this complete unbalanced dataset by considering an additional assumption that the repeated responses may further be influenced by an individual random effect. They have, thus, used the conditionally serially correlated model (8.1) to fit the data, whereas the model used in Section 6.7 was serially correlated without any attention to the individual random effects.

Note that even though we consider the complete data record from 180 individuals, it was found that four individuals had visibly distinct outlier responses. To understand the nature of the bulk of the data, for convenience, we have ignored these four responses and analyzed the data by using a possible nonstationary (due to age) correlation structure based longitudinal Poisson mixed model. Note that as opposed to the longitudinal fixed model (Chapter 6), the fitting of a longitudinal mixed model to this data set appears to be more reasonable. This is because as shown in Table 8.7, the average variation (for 176 individuals) of longitudinal responses was found to be 7.417 as opposed to the average sample mean 3.932. Thus the data appear to exhibit some overdispersion which motivated us to use the longitudinal Poisson mixed model instead of longitudinal Poisson fixed model.

Table 8.7 Observed and estimated summary statistics for health data for 176 individuals under both longitudinal mixed and longitudinal fixed models.

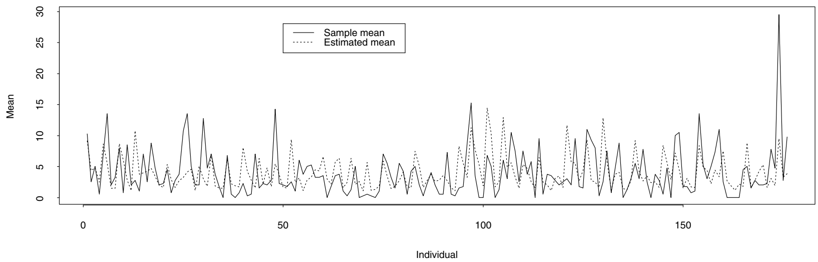
Quantity	Sample	Estimated Based on LM model	Estimated based on LF model
Mean	3.932	3.966	3.690
Variance	7.417	7.674	3.690

Table 8.8 Estimates of regression and variance component parameters with their estimated standard errors, as well as estimates of autocorrelations, under both longitudinal mixed and longitudinal fixed models for health data.

Model	Quantity	Gender	No. of Chronic Disease	Education	Age	Variance Component(σ^2)	Correlation Parameter(ρ)
LM	Estimate	−0.481	0.571	0.350	0.024	0.150	0.294
	ESE	0.069	0.077	0.065	0.001	0.076	–
LF	Estimate	−0.383	0.562	0.302	0.023	–	0.575
	ESE	0.051	0.060	0.047	0.001	–	–

Note that in the present analysis, the education level was coded in a reverse way as compared to that of Section 6.7 [see also Sutradhar (2003)], whereas the other covariates were coded in the same way. By applying the GQL estimating equation (8.18) for unbalanced data, we have obtained the estimates of the effects of gender ($\hat{\beta}_1$), chronic disease status ($\hat{\beta}_2$), education level ($\hat{\beta}_3$), and age ($\hat{\beta}_4$). Next, we have used the 'working' GQL estimating equation (8.44) for unbalanced data for the estimation of the random effects variance component σ_γ^2. Also, the longitudinal correlation parameter (ρ) was estimated by the moment equation (8.54) using formulas from (8.55) for the unbalanced data. The estimates along with their standard errors are given in Table 8.8.

Furthermore, the GQL approach was applied to estimate the regression effects and longitudinal correlation parameter by ignoring the presence of overdispersion. The results under this longitudinal fixed model (shown in the same Table 8.8) appear to agree quite well with those of Section 6.7. Note, however, that when the estimates under both a mixed and a fixed model were used to compute the fitted mean and variance of the data (Table 8.7), the results based on the mixed model appear to agree quite well with the sample mean and variance as compared to those of the fixed model. Because of this good agreement, in Figure 8.1, we exhibit the plots for the means and variances of longitudinal values of 176 individuals computed (1) from the sample as well as (2) from the fitted values based on the mixed model. The sample means and variances appear to agree quite well with the corresponding means and variances estimated based on the mixed model.



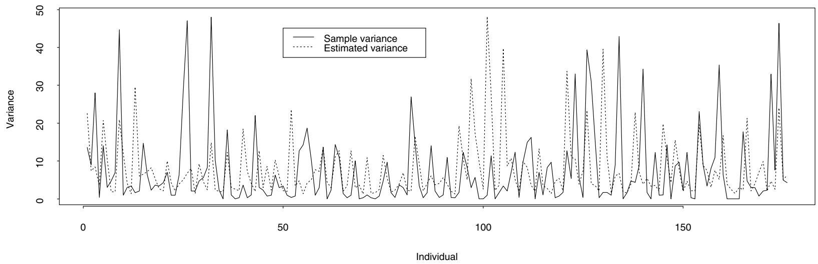


Fig. 8.1 Average and variance of $T = 4$ longitudinal counts and corresponding estimates for 176 individuals.

Recall from the simulation study (Table 8.6) that when σ_γ^2 was ignored but the data were generated following the full model, the longitudinal correlation was in general overestimated. The results in Table 8.8 appear to follow this pattern. This is because under the fixed model, $\hat{\rho}$ was found to be 0.575 as compared to $\hat{\rho} = 0.294$ under the mixed model. This also indicates that it is better to use the mixed model based results for this dataset as compared to those of the fixed model in Section 6.7.

8.3 A Mean Deflated Conditional Serially Correlated Model

When it is assumed that

$$\gamma_i \overset{i.i.d.}{\sim} N(0, \sigma_\gamma^2),$$

one may still consider that the conditional model (8.1) holds for the repeated responses, but with a deflated marginal mean such that $y_{it} \sim \text{Poi}(\mu_{it}^*)$ with

$$\mu_{it}^* = E[Y_{it}|\gamma_i] = \exp[x_{it}'\beta - \frac{1}{2}\sigma_\gamma^2 + \gamma_i], \tag{8.59}$$

yielding the unconditional mean as

$$\mu_{it} = E[Y_{it}] = \exp(x'_{it}\beta). \tag{8.60}$$

Note that this unconditional mean is free from σ_γ^2, whereas the unconditional mean under the model (8.1) is $\mu_{it} = \exp(x'_{it}\beta + \frac{1}{2}\sigma_\gamma^2)$, as given in (8.5). The formulas for the elements of $\Sigma_i = \text{cov}[Y_i]$ remain the same as in (8.6) and (8.8), except that μ_{it} for these formulas is now defined by (8.60).

8.3.1 First− and Second-Order Raw Response Based GQL Estimation

Under the present mean deflated model, the GQL estimating equation (8.18) for β takes the form

$$\sum_{i=1}^{K} \frac{\partial \mu_i'(\beta)}{\partial \beta} \Sigma_i^{-1}(\beta, \sigma_\gamma^2, \rho)[y_i - \mu_i(\beta)] = 0, \tag{8.61}$$

where the $\mu_i(\cdot)$ vector is free from σ_γ^2. Also, the ρ parameter is estimated by the method of moments using the formula in (8.54) except that $\mu_{it} = \exp(x'_{it}\beta)$ under the mean deflation model.

For the estimation of σ_γ^2, one may, however, use one of the following two 'working' GQL approaches.

8.3.1.1 GQL(I) Approach for σ_γ^2 Estimation

The formulas for the 'working' independence assumption based GQL estimating equation for σ_γ^2 remains the same as in (8.44), except that $\mu_{it} = \exp(x'_{it}\beta)$ is now free from σ_γ^2. Note that this new formula for the mean function μ_{it} has to be taken into account when derivatives are computed for (8.44) to derive the estimating equation for σ_γ^2. To make even more clear, the GQL(I) represents the GQL approach where σ_γ^2 is solved by using the 'working' independence (I) assumption based estimating equation, whereas the regression and the longitudinal correlation parameters are estimated by using the exact GQL and method of moments, respectively.

In the next approach, σ_γ^2 is computed by solving a GQL estimating equation with weight matrix computed based on a 'working' normality (N) assumption for the count responses. This approach is referred to as the GQL(N) approach.

8.3.1.2 GQL(N) Approach for σ_γ^2 Estimation

In this approach, β is estimated by using the GQL estimating equation (8.61), and the ρ parameter is estimated by the method of moments using $\mu_{it} = \exp(x'_{it}\beta)$ in (8.54). However, for the estimation of σ_γ^2, as opposed to solving the independence assumption based 'working' GQL estimating equation (8.44), we now solve the normality based 'working' GQL estimating equation

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta, \sigma_\gamma^2, \rho)}{\partial \sigma_\gamma^2} \Omega_{iN}^{-1}(\beta, \sigma_\gamma^2, \rho)[u_i - \lambda_i(\beta, \sigma_\gamma^2, \rho)] = 0, \tag{8.62}$$

where for

$$u_i = (y_{i1}^2, \ldots, y_{iT}^2, y_{i1}y_{i2}, \ldots, y_{it}y_{i,t+1}, \ldots, y_{i,T-1}y_{iT})',$$

one writes

$$\lambda_i(\beta, \sigma_\gamma^2, \rho) = E[U_i] = (\lambda_{i11}, \ldots, \lambda_{itt}, \ldots, \lambda_{iTT}, \lambda_{i12}, \ldots, \lambda_{iut}, \ldots, \lambda_{i,T-1,T})',$$

with

$$\lambda_{iut} = E[Y_{iu}Y_{it}] = \begin{cases} \mu_{it} + [\exp(\sigma_\gamma^2)]\mu_{it}^2, & \text{for } u = t \\ \rho^{t-u}\mu_{iu} + [\exp(\sigma_\gamma^2)]\mu_{iu}\mu_{it} & \text{for } u < t, \end{cases} \tag{8.63}$$

where $\mu_{it} = x_{it}'\beta$.

To compute the $\Omega_{i,N}^{-1}$ matrix for (8.62), we pretend that even though the response vector $y_i = (y_{i1}, \ldots, y_{iT})'$ of the repeated counts for the ith firm/individual are generated by AR(1) type count data model [(8.1 and 8.59)], they follow a 'working' multivariate normal distribution $N(\mu_i, \Sigma_i)$, where μ_i and Σ_i are the true mean vector and covariance matrix of y_i as in (8.61). Note that for the vector of count responses y_i, the elements of the μ_i vector are given by (8.60) and the diagonal and off-diagonal elements of the Σ_i matrix are given by

$$\sigma_{iut} = \lambda_{iut} - \mu_{iu}\mu_{it}, \text{ for } u \le t,$$

where λ_{iut} is given in (8.63). Next, these variances and covariances are used to compute the product moments of order four $E[Y_{iu}Y_{it}Y_{i\ell}Y_{im}]$, under the normality assumption. These product moments are then used to compute the 'working' fourth-order moments matrix $\Omega_{i,N}$.

Note that under normality,

$$E[(Y_{iu} - \mu_{iu})(Y_{it} - \mu_{it})(Y_{i\ell} - \mu_{i\ell})(Y_{im} - \mu_{im})] = \sigma_{iut}\sigma_{i\ell m} + \sigma_{iu\ell}\sigma_{itm} + \sigma_{ium}\sigma_{it\ell}. \tag{8.64}$$

Let

$$\delta_{iut\ell} = E[Y_{iu}Y_{it}Y_{i\ell}] \tag{8.65}$$

be the third-order product moments. By using (8.64), one may then obtain the fourth-order product moments under the normality as

$$\begin{aligned} E[Y_{iu}Y_{it}Y_{i\ell}Y_{im}] = & \ \sigma_{iut}\sigma_{i\ell m} + \sigma_{iu\ell}\sigma_{itm} + \sigma_{ium}\sigma_{it\ell} \\ & + \delta_{iut\ell}\mu_{im} + \delta_{iutm}\mu_{i\ell} + \delta_{iu\ell m}\mu_{it} + \delta_{it\ell m}\mu_{iu} \\ & - \sigma_{iut}\mu_{i\ell}\mu_{im} - \sigma_{iu\ell}\mu_{it}\mu_{im} - \sigma_{ium}\mu_{it}\mu_{i\ell} - \sigma_{it\ell}\mu_{iu}\mu_{im} \\ & - \sigma_{itm}\mu_{iu}\mu_{i\ell} - \sigma_{i\ell m}\mu_{iu}\mu_{it} + 3\mu_{iu}\mu_{it}\mu_{i\ell}\mu_{im}. \end{aligned} \tag{8.66}$$

As far as the third-order product moments are concerned, the normality based equation

$$E[(Y_{iu} - \mu_{iu})(Y_{it} - \mu_{it})(Y_{i\ell} - \mu_{i\ell})] = 0, \tag{8.67}$$

yields them as

$$\delta_{iut\ell} = \sigma_{iut}\mu_{i\ell} + \sigma_{iu\ell}\mu_{it} + \sigma_{it\ell}\mu_{iu} - 2\mu_{iu}\mu_{it}\mu_{i\ell}. \tag{8.68}$$

8.3.2 Corrected Response (CR) Based GQL Estimation

Note that the computation of the Ω_{iN} by using (8.66) and (8.68) is straightforward but lengthy. To reduce the computational burden, one may use a corrected (from mean) second-order response vector based 'working' GQL approach for the estimation of σ_γ^2. The regression effects β and the longitudinal correlation parameter ρ, are still estimated by solving (8.61) and (8.54), respectively.

Consider a $\{T(T+1)/2\}$-dimensional vector statistic

$$\begin{aligned} g_i^* = {} & [(y_{i1} - \mu_{i1})^2, \ldots, (y_{iT} - \mu_{iT})^2, (y_{i1} - \mu_{i1})(y_{i2} - \mu_{i2}), \\ & \ldots, (y_{i(T-1)} - \mu_{i(T-1)})(y_{iT} - \mu_{iT})]', \end{aligned} \tag{8.69}$$

based on corrected squared and pairwise products. Let $\tilde{\sigma}_i$ be the mean of g_i^*. Note that because Σ_i, the covariance matrix of the response vector y_i, is known by (8.60) and (8.61), the elements of the $\tilde{\sigma}_i$ vector are nothing but the selected elements of the Σ_i matrix. In notation,

$$\tilde{\sigma}_i = [\sigma_{i11}, \ldots, \sigma_{iTT}, \sigma_{i12}, \ldots, \sigma_{i(T-1)T}]', \tag{8.70}$$

where

$$\sigma_{itt} = \mu_{it} + c\mu_{it}^2,$$

and for $t < w$,

$$\sigma_{itw} \equiv \sigma_{iwt} = \rho^{w-t}\mu_{it} + c\mu_{it}\mu_{iw},$$

with $\mu_{it} = \exp(x_{it}'\beta)$, and $c = [\exp(\sigma_\gamma^2) - 1]$. Suppose that the covariance matrix of g_i^* is denoted by $\Omega_i^*(\beta,\, \sigma_\gamma^2,\, \rho)$.

8.3.2.1 GQL(CR-I) Estimation for σ_γ^2

By using the distance vector $g_i^* - \tilde{\sigma}_i$, in the manner similar to that of the GQL(I) approach [see also (8.44)], one may construct the 'working' GQL estimating equation for σ_γ^2 as

$$\sum_{i=1}^{K} \frac{\partial \tilde{\sigma}_i'}{\partial \sigma_\gamma^2} {\Omega^*}_i^{-1}(\beta,\, \sigma_\gamma^2,\, \rho = 0)(g_i^* - \tilde{\sigma}_i) = 0, \tag{8.71}$$

where

$$\frac{\partial \tilde{\sigma}_i'}{\partial \sigma_\gamma^2} = [\tilde{\sigma}_{i11}, \ldots, \tilde{\sigma}_{iTT}, \tilde{\sigma}_{i12}, \ldots, \tilde{\sigma}_{i(T-1)T}]'$$

$$= \exp(\sigma_\gamma^2)[\mu_{i1}^2, \ldots, \mu_{iT}^2, \mu_{i1}\mu_{i2}, \ldots, \mu_{i,T-1}\mu_{iT}]'$$

$$= \tilde{v}_i' \text{ (say)}. \quad (8.72)$$

Let $\hat{\sigma}^2_{\gamma,GQL(CR-I)}$ be the solution of (8.71) for σ^2_γ. By abbreviating $\Omega_i^*(\beta, \sigma^2_\gamma, \rho = 0)$ with $\Omega_i^*(I)$, this estimate may be obtained by using the iterative equation

$$\hat{\sigma}^2_{\gamma,(m+1)} = \hat{\sigma}^2_{\gamma,(m)} + \left[\sum_{i=1}^K \tilde{v}_i' {\Omega^*}_i^{-1}(I)\tilde{v}_i\right]_m^{-1} \left[\sum_{i=1}^K \tilde{v}_i' {\Omega^*}_i^{-1}(I)(g_i^* - \tilde{\sigma}_i)\right]_m. \quad (8.73)$$

Furthermore, one may show [see also (8.49)] that $K^{1/2}(\hat{\sigma}^2_{\gamma,GQL(CR-I)} - \sigma^2_\gamma)$ has a univariate normal distribution, as $K \to \infty$, with mean zero, and the variance which can be consistently estimated by

$$K\left[\sum_{i=1}^K \tilde{v}_i' {\Omega^*}_i^{-1}(I)\tilde{v}_i\right]^{-2} \left[\sum_{i=1}^K \tilde{v}_i' {\Omega^*}_i^{-1}(I)(g_i^* - \tilde{\sigma}_i)(g_i^* - \tilde{\sigma}_i)' {\Omega^*}_i^{-1}(I)\tilde{v}_i\right]. \quad (8.74)$$

Construction of $\Omega_i^*(I)$

As far as the formulas for the elements of the $\Omega_i^*(I)$ matrix are concerned, we provide them through Lemma 8.1 and Exercise 8.1.

Lemma 8.1 The formulas for $\text{var}[(Y_{it} - \mu_{it})^2]$ and $\text{cov}[(Y_{it} - \mu_{it})^2, (Y_{it} - \mu_{it})(Y_{is} - \mu_{is})]$ are given by

$$\text{var}[(Y_{it} - \mu_{it})^2] = \mu_{it} + \{7\exp(\sigma^2_\gamma) - 4\}\mu_{it}^2 + 6\{\exp(3\sigma^2_\gamma) - 2\exp(\sigma^2_\gamma) + 1\}\mu_{it}^3$$

$$+\{\exp(6\sigma^2_\gamma) - 4\exp(3\sigma^2_\gamma) + 6\exp(\sigma^2_\gamma) - 3\}\mu_{it}^4 - \sigma^2_{itt}, \quad (8.75)$$

and

$$\text{cov}[(Y_{it} - \mu_{it})^2 \ , \ (Y_{it} - \mu_{it})(Y_{is} - \mu_{is})] = \mu_{it}\mu_{is}\left[\{\exp(\sigma^2_\gamma) - 1\} + 3\{\exp(3\sigma^2_\gamma)\right.$$

$$\left. -2\exp(\sigma^2_\gamma) + 1\}\mu_{it}\right] - \sigma_{itt}\sigma_{its}(0), \quad (8.76)$$

respectively, where σ_{itt} is the variance of y_{it} as given by (8.70), and $\sigma_{its}(0) = c\mu_{it}\mu_{is}$ is the covariance between y_{it} and y_{is} evaluated from (8.70) at $\rho = 0$.

Proof of Lemma 8.1: To derive the formula in (8.75), we re-express the variance as

$$\text{var}[(Y_{it} - \mu_{it})^2] = E(Y_{it} - \mu_{it})^4 - \{E(Y_{it} - \mu_{it})^2\}^2$$

$$= E_{\gamma_i}E[\{(Y_{it} - \mu_{it}^*) + (\mu_{it}^* - \mu_{it})\}^4 | \gamma_i] - \sigma^2_{itt}$$

$$= E_{\gamma_i} E[\{Z_{it}^4 + 4Z_{it}^3 b_{it} + 6Z_{it}^2 b_{it}^2 + 4Z_{it} b_{it}^3 + b_{it}^4\}|\gamma_i] - \sigma_{itt}^2, \quad (8.77)$$

with $Z_{it} = Y_{it} - \mu_{it}^*$ and $b_{it} = \mu_{it}^* - \mu_{it}$. The result in (8.75) may now be obtained from (8.77) by noting the fact that conditional on γ_i, Y_{it} has the Poisson distribution with mean parameter μ_{it}^*; that is,

$$E(Z_{it}|\gamma_i) = 0,\ E(Z_{it}^2|\gamma_i) = E(Z_{it}^3|\gamma_i) = \mu_{it}^*,\ \text{and}\ E(Z_{it}^4|\gamma_i) = \mu_{it}^* + 3\mu^{*2}_{it},$$

and

$$b_{it} = \mu_{it}^* - \mu_{it} = \exp(x_{it}'\beta - \frac{1}{2}\sigma_\gamma^2)\{w_i - \exp(\frac{1}{2}\sigma_\gamma^2)\},$$

where

$$E\{W_i - \exp(\frac{1}{2}\sigma_\gamma^2)\}^2 = \exp(\sigma_\gamma^2)\{\exp(\sigma_\gamma^2) - 1\}, \quad (8.78)$$

$$E\{W_i - \exp(\frac{1}{2}\sigma_\gamma^2)\}^3 = \exp(\frac{9}{2}\sigma_\gamma^2) + 2\exp(\frac{3}{2}\sigma_\gamma^2) - 3\exp(\frac{5}{2}\sigma_\gamma^2), \quad (8.79)$$

$$E\{W_i - \exp(\frac{1}{2}\sigma_\gamma^2)\}^4 = \exp(8\sigma_\gamma^2) + 6\exp(3\sigma_\gamma^2) - 4\exp(5\sigma_\gamma^2) - 3\exp(2\sigma_\gamma^2). \quad (8.80)$$

The formula for the covariance in (8.76) may be obtained similarly.

8.3.2.2 GQL(CR-N) Estimation σ_γ^2

Under this approximation, instead of the independence assumption based estimating equation (8.71), one solves the normality assumption based estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \tilde{\sigma}_i'}{\partial \sigma_\gamma^2} \Omega^{*-1}_{i,N}(\beta, \sigma_\gamma^2, \rho)(g_i^* - \tilde{\sigma}_i) = 0, \quad (8.81)$$

to obtain a 'working' GQL estimate for σ_γ^2. Because g_i^* in (8.81) is a vector of corrected squares and cross-products of repeated count responses as in (8.69), the fourth-order moments matrix $\Omega_{i,N}^*(\beta, \sigma_\gamma^2, \rho)$ in (8.71), under normality assumption, may be computed simply by applying the equation (8.64); that is,

$$E[(Y_{iu} - \mu_{iu})(Y_{it} - \mu_{it})(Y_{i\ell} - \mu_{i\ell})(Y_{im} - \mu_{im})] = \sigma_{iut}\sigma_{i\ell m} + \sigma_{iu\ell}\sigma_{itm} + \sigma_{ium}\sigma_{it\ell},$$

where σ_{itt}, for example, by (8.63), has the formula

$$\sigma_{itt} = \mu_{it} + [\exp(\sigma_\gamma^2) - 1]\mu_{it}^2,$$

with $\mu_{it} = \exp(x_{it}'\beta)$.

8.3.3 Relative Performances of GQL(I) and GQL(N) Estimation Approaches: A Simulation Study

A simulation study conducted in Section 8.2.4 suggested that the GQL(I) approach produces almost unbiased estimates for the regression (β) and longitudinal correlation (ρ) parameters, but it underestimated the overdispersion parameter σ_γ^2, especially when the true value of σ_γ^2 is large. We remark that this underestimation happens because of the approximation required to construct the GQL estimating equation in the longitudinal mixed model setup, whereas the GQL approach in the familial mixed model setup does not require any such approximations and yields an almost unbiased estimate for this overdispersion parameter (see Table 4.6, Section 4.2.7, for example). In this section, we examine whether the normality approximation based GQL(N) approach improves the estimation, mainly for σ_γ^2. This we do, however, under the present mean deflated model. As far as the simulation design is concerned, we use the same balanced design D_1 as used in the simulation study in Section 8.2.4.

For the sake of completeness, we use all four approximations, namely the GQL(I), GQL(N), GQL(CR-I), and GQL(CR-N) approaches, and estimate all three parameters β, ρ, and σ_γ^2.

8.3.3.1 Performance for Overdispersion Estimation

The simulated mean, standard error, and mean squared error are reported in Table 8.9 for the overdispersion parameter σ_γ^2 under all four approximations.

The results of Table 8.9 suggest that GQL(CR-I) approximation provides an estimate of σ_γ^2 with smaller MSE as compared to the GQL(I) approximation. Similarly, between the normality based approximations GQL(N) and GQL(CR-N), the GQL(CR-N) approximation produces the σ_γ^2 estimate with smaller MSE as compared to the GQL(N) approximation.

For example, when $\rho = 0.75$, the GQL(I) approximation based approach estimates $\sigma_\gamma^2 = 0.75$ with MSE 0.067, whereas the GQL(CR-I) approach estimates this parameter value with smaller MSE 0.053. Similarly, the same parameter value is estimated by GQL(N) and GQL(CR-N) approximations with MSE 0.128 and 0.041, respectively. Thus the corrected response based approximations appear to perform better as compared to their corresponding raw response based approximations. This pattern appears to hold for all parameter values considered in Table 8.9. Among all four approximations, the GQL(CR-N) approximation performs the best and GQL(N) approximation is the worst. Note that the GQL(CR-N) approximation attains the smallest MSE because of the smallest standard error of the estimate. Thus the efficiency gain by the GQL(CR-N) approximation may be quite significant as compared to the other three approximations in estimating the overdispersion parameter σ_γ^2. In summary, normality based approximation appears to work well when CRs are used to construct the GQL estimating equation. If raw or uncorrected responses are used, independence assumption based approximation works better than the normality based approximation. Further note that irrespective of the approxima-

Table 8.9 Overdispersion estimation: Simulated mean, simulated standard error, and simulated mean squared errors of the GQL estimates for the overdispersion parameter σ_γ^2, based on GQL(I), GQL(CR-I), GQL(N), and GQL(CR-N) approximations, for selected values of σ_γ^2 and ρ; $K = 100$; $T = 4$; $\beta_1 = \beta_2 = 0.0$; 1000 simulations.

			Approximation			
True σ_γ^2	True ρ	Quantity	GQL(I)	GQL(CR-I)	GQL(N)	GQL(CR-N)
0.25	0.25	SM	0.247	0.248	0.377	0.235
		SSE	0.148	0.126	0.267	0.079
		SMSE	0.022	0.016	0.087	0.006
	0.50	SM	0.247	0.247	0.408	0.237
		SSE	0.171	0.130	0.340	0.103
		SMSE	0.029	0.017	0.141	0.011
	0.75	SM	0.296	0.281	0.371	0.242
		SSE	0.193	0.172	0.280	0.135
		SMSE	0.039	0.030	0.093	0.018
0.50	0.25	SM	0.457	0.457	0.470	0.456
		SSE	0.187	0.163	0.211	0.106
		SMSE	0.037	0.028	0.045	0.013
	0.50	SM	0.456	0.458	0.525	0.452
		SSE	0.205	0.179	0.292	0.128
		SMSE	0.044	0.034	0.086	0.019
	0.75	SM	0.594	0.572	0.595	0.507
		SSE	0.239	0.203	0.327	0.175
		SMSE	0.066	0.046	0.116	0.031
0.75	0.25	SM	0.771	0.762	0.802	0.754
		SSE	0.212	0.180	0.194	0.133
		SMSE	0.045	0.032	0.040	0.018
	0.50	SM	0.795	0.785	0.799	0.742
		SSE	0.227	0.198	0.281	0.158
		SMSE	0.053	0.040	0.081	0.025
	0.75	SM	0.806	0.790	0.847	0.746
		SSE	0.253	0.226	0.344	0.203
		SMSE	0.067	0.053	0.128	0.041

tions, the MSE of the estimator of σ_γ^2 gets larger as the value of ρ increases. Thus the estimation of σ_γ^2 is affected by the longitudinal correlations. The performances of all four approximations in estimating the longitudinal correlation index parameter are discussed in Section 8.3.3.3.

8.3.3.2 Performance for Regression Effects Estimation

The GQL estimates of β_1 and β_2 are obtained by solving the GQL estimating equation (8.61). These estimates are affected by the other two parameters σ_γ^2 and ρ only through the weight matrix involved in (8.61). Thus it is expected that the estimates of these later parameters will have no effect on the consistency of the regression estimates. Furthermore, because the GQL(CR-I) and GQL(I) produced almost the same estimates for σ_γ^2 in Table 8.9, for convenience we exclude the tabulation of the results for the GQL(CR-I) approach, while showing results on regression estimates in Table 8.10. The simulation results, namely SMs, SSEs, and SMSEs reported in Table 8.10 show that the efficiencies of the GQL regression estimates are also not affected by the estimates of the σ_γ^2 as well as ρ parameters. This is because the MSEs of the estimates of β_1 and β_2 appear to remain almost the same under all three remaining approximations.

Table 8.10 Regression estimation: Simulated mean, simulated standard error, and simulated mean squared errors of the GQL estimates for the regression parameters β_1 and β_2, based on GQL(I), GQL(N), and GQL(CR-N) approximations, for selected values of σ_γ^2 and ρ; $K = 100$; $T = 4$; $\beta_1 = \beta_2 = 0.0$; 1000 simulations.

			Approximation					
			GQL(I)		GQL(N)		GQL(CR-N)	
True σ_γ^2	True ρ	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.25	0.25	SM	0.043	−0.021	0.043	−0.024	0.043	−0.022
		SSE	0.158	0.084	0.157	0.084	0.158	0.085
		SMSE	0.027	0.007	0.026	0.008	0.027	0.008
	0.50	SM	0.014	−0.013	0.012	−0.014	0.036	−0.020
		SSE	0.126	0.079	0.126	0.079	0.130	0.078
		SMSE	0.016	0.006	0.016	0.006	0.018	0.006
	0.75	SM	0.017	−0.012	0.015	−0.012	0.018	−0.010
		SSE	0.100	0.064	0.101	0.165	0.102	0.066
		SMSE	0.010	0.004	0.010	0.027	0.011	0.004
0.50	0.25	SM	0.050	−0.027	0.050	−0.029	0.049	−0.025
		SSE	0.166	0.093	0.165	0.093	0.166	0.094
		SMSE	0.030	0.009	0.030	0.009	0.030	0.009
	0.50	SM	0.028	−0.016	0.030	−0.018	0.027	−0.016
		SSE	0.134	0.085	0.134	0.085	0.135	0.085
		SMSE	0.019	0.007	0.019	0.008	0.019	0.007
	0.75	SM	0.002	−0.006	0.002	−0.007	0.002	−0.006
		SSE	0.096	0.066	0.096	0.066	0.096	0.066
		SMSE	0.009	0.004	0.009	0.004	0.009	0.004
0.75	0.25	SM	0.016	−0.039	0.014	−0.037	0.014	−0.036
		SSE	0.154	0.090	0.153	0.090	0.153	0.090
		SMSE	0.024	0.010	0.024	0.009	0.024	0.009
	0.50	SM	0.017	−0.027	0.015	−0.026	0.014	−0.025
		SSE	0.131	0.085	0.130	0.085	0.130	0.085
		SMSE	0.017	0.008	0.017	0.008	0.017	0.008
	0.75	SM	0.009	−0.014	0.009	−0.015	0.009	−0.014
		SSE	0.098	0.067	0.098	0.067	0.098	0.067
		SMSE	0.010	0.005	0.010	0.005	0.010	0.005

For example, when $\sigma_\gamma^2 = 0.75$ and $\rho = 0.50$, all three approximations yield MSEs 0.017 for $\hat{\beta}_1$ and 0.008 for $\hat{\beta}_2$. Further note that under any given approximation, the MSEs of both $\hat{\beta}_1$ and $\hat{\beta}_2$ appear to get smaller as the value of ρ gets larger. For example, when $\sigma_\gamma^2 = 0.50$, under the GQL(CR-N) approximation, the MSEs of $\hat{\beta}_2$ are found to be 0.009, 0.007, and 0.004, for $\rho = 0.25,\ 0.50,$ and 0.75, respectively.

Furthermore, these MSE values appear to remain almost the same for the other two values of $\sigma_\gamma^2 = 0.25$ and 0.75. Thus, it is clear that the GQL estimates of the components of β are more affected by ρ than σ_γ^2. But, it has nothing to do with the selection of an approximation for β estimation. More specifically, any of the four approximations can be chosen for β estimation. We, however, recommend the GQL(CR-N) approximation as it is simpler and estimates σ_γ^2 with smaller MSEs as compared to the other three approximations.

8.3.3.3 Performance for Correlation Index Estimation

The longitudinal correlation index parameter ρ has to be estimated to understand the correlation structure of the data as well as for the estimation of β and σ_γ^2 parameters. This parameter may be consistently estimated by using the moment equation (8.54) using $\mu_{it} = \exp(x_{it}'\beta)$ under the mean deflated model. For selected true values of σ_γ^2, the simulated moment estimates of ρ as well as their SSEs and MSEs are reported in Table 8.11. The results of Table 8.11 show that the moments based approach produces almost unbiased estimates for ρ, irrespective of the approximations used to estimate σ_γ^2.

For example, when $\sigma_\gamma^2 = 0.75$, the correlation index parameter value $\rho = 0.5$ is estimated as 0.488, 0.469, 0.465, and 0.477, by using σ_γ^2 estimates based on the GQL(I), GQL(CR-I), GQL(N), and GQL(CR-N) approximations, respectively. However, the corresponding SMSEs were found to be 0.014, 0.012, 0.016, and 0.008, showing that the GQL(CR-N) approach produces slightly more efficient ρ estimates as compared to the other three approximations. This pattern appears to hold for other selected values of σ_γ^2 and ρ. Thus, the GQL(CR-N) approximation produces better or the same estimates for both σ_γ^2 (Table 8.9) and ρ (Table 8.11). Furthermore, this approach is as good as any other approximations in estimating β, which makes the GQL(CR-N) approximation best in estimating all parameters of the longitudinal mixed model.

8.3.4 A Further Application: Analyzing Patent Count Data

To illustrate the mean deflated longitudinal mixed model for count data introduced in Sections 8.3.1 and 8.3.2, we consider a part of the U.S. patents and R&D (Research and Development) expenditures dataset that contains the patents and R&D expenditures from 168 firms from 1971 to 1979. This patent count dataset was earlier analyzed by some authors such as Hausman, Hall, and Griliches (1984), Blundell, Griffith, and Windmeijer (1995), and Montalvo (1997), mainly by using the GMM and CML approaches. The patent data also contain the type of each firm whether scientific or nonscientific, and the log of the book value of capital (in 1972 millions of dollars) less than or equal to 4.0. These two covariates along with the R&D expenditures for the period from 1971 to 1979 are shown in Table 8B in the appendix. As far as the longitudinal responses are concerned, we consider the patent

Table 8.11 Correlation estimation: Simulated mean, simulated standard error, and simulated mean squared error of the moment estimates for correlation index parameter ρ, based on GQL(I), GQL(I), GQL(N), and GQL(CR-N) approximations, for selected values of σ_γ^2 and ρ; $K = 100$; $T = 4$; $\beta_1 = \beta_2 = 0.0$; 1000 simulations.

			Approximations			
True σ_γ^2	True ρ	Quantity	GQL(I)	GQL(CR-I)	GQL(N)	GQL(CR-N)
0.25	0.25	SM	0.269	0.261	0.258	0.263
		SE	0.096	0.086	0.110	0.075
		SMSE	0.010	0.008	0.012	0.006
	0.50	SM	0.466	0.460	0.490	0.505
		SE	0.081	0.075	0.082	0.063
		SMSE	0.008	0.007	0.007	0.006
	0.75	SM	0.748	0.748	0.749	0.751
		SE	0.052	0.047	0.053	0.043
		SMSE	0.003	0.002	0.003	0.002
0.50	0.25	SM	0.266	0.259	0.265	0.260
		SE	0.114	0.105	0.120	0.091
		SMSE	0.013	0.011	0.015	0.008
	0.50	SM	0.509	0.513	0.524	0.505
		SE	0.097	0.084	0.107	0.077
		SMSE	0.009	0.007	0.012	0.006
	0.75	SM	0.728	0.737	0.735	0.739
		SE	0.065	0.058	0.058	0.050
		SMSE	0.005	0.004	0.004	0.003
0.75	0.25	SM	0.243	0.245	0.174	0.218
		SE	0.159	0.131	0.164	0.118
		SMSE	0.025	0.017	0.033	0.015
	0.50	SM	0.488	0.469	0.465	0.477
		SE	0.118	0.104	0.121	0.090
		SMSE	0.014	0.012	0.016	0.008
	0.75	SM	0.732	0.740	0.739	0.734
		SE	0.073	0.065	0.078	0.059
		SMSE	0.006	0.004	0.006	0.004

counts from 1974 to 1979 awarded to each of 168 industries. These patent counts are shown in Table 8A in the appendix.

In the notation of Sections 8.3.1 and 8.3.2, suppose that y_{it} is the number of patents for the ith firm at time t. Thus, for $t = 1,\ldots,T$, with $T = 6$, y_{i1} denotes the number of patents awarded to the ith firm in 1974, and y_{i2} denotes the number of patents awarded in 1975, and so on. As far as the explanatory variables are concerned, we consider $p = 6$ covariates, among which $p_1 = 4$ are time-dependent covariates and $p_2 = 2$ are time-independent covariates, so that $x_{it} = (x_{it1},\ldots,x_{it6})'$, where for $u = 1,\ldots,p$, x_{itu} denotes the uth covariate value recorded at year t from the ith $(i = 1,\ldots,K)$ firm for $K = 168$. To be specific, x_{it1}, x_{it2}, x_{it3}, and x_{it4} are the R&D expenditures at year t, $t-1$, $t-2$, and $t-3$, respectively. Similarly, x_{it5} denotes the type of firm (coded as 0 for nonscientific firms and as 1 for scientific firms) and x_{it6} is the log of the book value of capital in 1972. In this problem, it is of main interest to find the relationship of the above six covariates (x_{it}) with the number of patents (y_{it}) awarded at each of the six years from 1974 to 1979. The count responses y_{it}, on top of x_{it}, may also be influenced by certain unobservable random effects (γ_i), and because the repeated count responses of a firm may be longitudinally correlated, the longitudinal mixed model (8.1) appears therefore to be appropriate to fit the data. As far as the conditional means are concerned, we choose to use the simpler mean deflated model given in (8.59). Thus, the estimation techniques introduced in Sections 8.3.1 and 8.3.2 would be appropriate to use to analyze the present patent count dataset.

Note that to see whether the model (8.60) and (8.63) can represent the patents and R&D data, we have computed the basic moments (mean, variance, correlations) of the data as shown in Table 8.12. It is clear that at a given year, the variances of the count response are much larger than their corresponding means.

Table 8.12 Six yearly summary statistics for 168 firms.

	1974	1975	1976	1977	1978	1979
Data based moments						
Mean	2.952	2.435	2.369	2.244	2.399	2.161
SD	3.508	2.952	2.704	2.851	3.094	2.854
Correlation	1.000	0.692	0.615	0.561	0.648	0.635
		1.000	0.651	0.607	0.583	0.534
			1.000	0.595	0.593	0.592
				1.000	0.630	0.661
					1.000	0.695
						1.000

For example, the mean of the number of patents awarded in 1977 is 2.244, whereas the variance is 8.128. This mean−variance relation may well be explained by the formulas for mean $\mu_{it} = \exp(x'_{it}\beta)$ from (8.60) and variance $\sigma_{itt} = \mu_{it} + [\exp(\sigma^2) - 1]\mu_{it}^2$ from (8.63), which are derived from the model (8.1) and (8.59). In fact the estimate of β, obtained by solving (8.61) iteratively, and the estimate of σ_γ^2 obtained by using all four approximations, namely GQL(I) [(8.44) with μ_{it} as in (8.60)], GQL(N) (8.62), GQL(CR-I) (8.71), and GQL (CR-N) (8.81) yielded the mean and standard deviations close to the observed means and standard deviations. For example, the estimated mean and standard deviations based on GQL(CR-I) and GQL(CR-N) approximations, shown in Table 8.13, are in good agreement with those observed means and standard deviations in Table 8.12. Note that in the simulation study conducted in the last section, these two approximations were found to be better than the other two approximations in estimating all parameters, GQL(CR-N) being the best.

Note that for the aforementioned estimation of β as discussed above, we also had to compute the moment estimate of ρ under each approximation. The estimates of all parameters along with standard errors, where applicable, based on the two best

Table 8.13 Yearly fitted values based on GQL(CR-I) and GQL(CR-N) approximations.

	1974	1975	1976	1977	1978	1979
GQL(CR-I) based fitted moments						
Mean	2.304	2.226	2.305	2.348	2.417	2.432
SD	2.266	2.213	2.278	2.321	2.374	2.388
Correlation	1.000	0.504	0.450	0.444	0.447	0.447
		1.000	0.501	0.446	0.443	0.442
			1.000	0.506	0.453	0.447
				1.000	0.509	0.455
					1.000	0.513
						1.000
GQL(CR-N) based fitted moments						
Mean	2.304	2.226	2.305	2.348	2.417	2.432
SD	2.341	2.286	2.362	2.400	2.455	2.470
Correlation	1.000	0.526	0.476	0.471	0.473	0.473
		1.000	0.523	0.472	0.470	0.469
			1.000	0.528	0.479	0.473
				1.000	0.532	0.480
					1.000	0.536
						1.000

approximations, namely GQL(CR-N) and GQL(CR-I), are reported in Table 8.14. Before we interpret these, mainly the estimates of the regression effects, we examine how these estimates explain the overall longitudinal correlations of the data.

Table 8.14 GQL(CR-I) and GQL(CR-N) estimates of regression and their estimated standard errors, as well as estimates of autocorrelations under Poisson longitudinal mixed model for the patents and R&D data.

	Estimation Approach			
	GQL(CR-I) Approximation		GQL(CR-N) Approximation	
Parameters	Estimate	SE	Estimate	SE
Lag 0 R&D (x_1)	0.447	0.086	0.446	0.086
Lag 1 R&D (x_2)	−0.119	0.098	−0.119	0.096
Lag 2 R&D (x_3)	−0.007	0.090	−0.007	0.093
Lag 3 R&D (x_4)	−0.027	0.076	−0.027	0.076
Firm type (x_5)	0.343	0.101	0.343	0.102
Log book (x_6)	0.265	0.083	0.265	0.026
σ_γ^2	0.370	0.074	0.336	0.008
ρ	0.452	–	0.496	–

For this purpose, we have first computed the observed correlations up to lag 5 (from six years' data) from 168 firms. These observed correlations are shown in Table 8.12 along with the observed means and standard deviations for six years.

Next, we estimate all lag correlations up to lag 5 by using the lag $\ell = 1, \ldots, T-1$, correlation formula

$$\text{corr}(Y_{it}, Y_{i,t-\ell}) = \frac{\rho^{\ell}\mu_{i,t-\ell} + c\mu_{it}\mu_{i,t-\ell}}{[\{\mu_{it} + c\mu_{it}^2\}\{\mu_{i,t-\ell} + c\mu_{i,t-\ell}^2\}]^2}, \tag{8.82}$$

constructed from (8.9) by using (8.60), where $c = \exp(\sigma_\gamma^2) - 1$. Now by averaging over the 168 firms, we compute the estimate of lag $\ell = 1, \ldots, T-1$, correlations under a given approximation. These estimated lag correlations under the GQL(CR-I) and GQL(CR-N) approximations are displayed in Table 8.13. Note that the estimated lag correlations shown in Table 8.13 are in general in agreement with the observed lag correlations displayed in Table 8.12. This agreement along with the agreement of the observed and estimated means and standard deviations of the data provides a satisfactory assessment about the use of the model (8.1) along with (8.60) to fit the patent and R&D data and about the performance of the estimation approximations.

We now interpret the estimates provided in Table 8.14 under both the GQL(CR-I) and GQL(CR-I) approaches. It is clear from the results of this table that both the GQL(CR-I) and GQL(CR-N) approaches yield the same estimates along with the same standard errors for the components of the β parameter. This is in agreement with the simulation findings displayed in Table 8.10 for the estimation of the β parameter. Based on the GQL(CR-I) approximation, the current (lag 0) R&D expenditures appear to have a positive influential effect (0.446) on the patents awarded to the firm, whereas the lag 1, 2, and 3 R&D covariates appear to have moderately negative effect or no effects on the patent numbers, estimates being −0.119, −0.007 and −0.027 respectively. The firm type appears to have a large positive effect (0.343) indicating that the scientific firms are awarded more patents as compared to the nonscientific firms. As far as the capital value of the firm is concerned, it appears that it also has large positive effect (0.265) on the number of patents awarded to the firms.

With regard to the estimates of the overdispersion parameter σ_γ^2, the GQL(CR-I) approach yielded the σ_γ^2 estimate as 0.370 with standard error 0.074, whereas the GQL(CR-N) approach gives the σ_γ^2 estimate as 0.336 with smaller standard error 0.008. These results are in agreement with the simulation results displayed in Table 8.9 for the estimation of σ_γ^2, where it was found that the GQL(CR-N) approximation performs the best in estimating the parameters. The estimate of the ρ parameter based on the best GQL(CR-N) approximation is found to be 0.496, which is significantly far away from zero correlation. Note that the GQL(CR-N) estimate 0.336 for σ_γ^2 is the reflection of a large overdispersion in the data. This is also verified from Tables 8.12 and 8.13, where it was shown that the estimated and/or observed variances are larger than the corresponding means.

8.4 Longitudinal Negative Binomial Fixed Model and Estimation of Parameters

In previous sections, the longitudinal count responses of an individual are assumed to be longitudinally correlated conditional on the individual random effect. This causes unconditional correlations to be functions of both longitudinal correlation index parameter (ρ) and the variance of the random effects (σ_γ^2). Jowaheer and Sutradhar (2002) have considered a different model where the marginal variance is only affected by the variance of the random effects and longitudinal correlations are simply functions of the correlation index parameter ρ. Consequently, this model of Jowaheer and Sutradhar (2002) is in fact a longitudinal fixed model. We discuss this model and its basic properties below. Also, the estimation of the parameters of the model is discussed.

Marginal Mixed Model:

In Section 8.1, it was assumed that each of the repeated count responses $y_{i1},\ldots,y_{iT}$ shares the common random effect γ_i, and conditional on γ_i, the responses are longitudinally correlated. As far as the influence of random effects is concerned, we now consider that these T responses are influenced by T different random effects, namely $\gamma_{i1},\ldots,\gamma_{it},\ldots,\gamma_{iT}$, respectively, and these T random effects are independent. Thus, in the spirit of the previous sections, the random effects are not causing any longitudinal correlations among the repeated responses. More specifically, these random effects will cause certain overdispersion marginally on each responses.

Suppose that conditional on γ_{it}, y_{it} has the Poisson distribution given by

$$f(y_{it}|\gamma_{it}) = \frac{1}{y_{it}!}\exp\{y_{it}\eta_{it} - \exp(\eta_{it})\},$$

[see (8.19)] with $E(Y_{it}|\gamma_{it}) = \text{var}(Y_{it}|\gamma_{it}) = \exp(\eta_{it})$, where $\eta_{it} = x_{it}^\top\beta + \log(\gamma_{it})$. Next suppose that γ_{it} has the gamma distribution with mean 1 and variance c^*, with density

$$g(\gamma_{it}) = \left\{c^{*-1/c^*}/\Gamma(c^*-1)\right\}\exp(-c^{*-1}\gamma_{it})\gamma_{it}^{c^{*-1}-1}. \tag{8.83}$$

For $\theta_{it} = x_{it}'\beta$, it then follows that marginally y_{it} has the negative binomial distribution given by

$$f(y_{it}) = \frac{\Gamma\left(c^{*-1}+y_{it}\right)}{\Gamma(c^{*-1})y_{it}!}\left(\frac{1}{1+c^*\theta_{it}}\right)^{c^{*-1}}\left(\frac{c^*\theta_{it}}{1+c^*\theta_{it}}\right)^{y_{it}}, \tag{8.84}$$

which accommodates the overdispersion indexed by c^*. More specifically, under (8.84), the marginal expectation and the variance have the formulas

$$E(Y_{it}) = \theta_{it} = \exp(x_{it}'\beta)$$

$$\text{var}(Y_{it}) = \theta_{it} + c^* \theta_{it}^2. \tag{8.85}$$

The negative binomial distribution (8.84) will be denoted by $y_{it} \sim \text{NeBi}(1/c^*, c^* \theta_{it})$.

8.4.1 Inferences in Stationary Negative Binomial Correlation Models

Following Lewis (1980) and McKenzie (1986), we may relate y_{it} for the ith individual at time t with $y_{i,t-1}$ by

$$y_{it} = \alpha_{it} * y_{i,t-1} + d_{it}, \tag{8.86}$$

where, for given probability $0 < \alpha_{it} < 1$ and count $y_{i,t-1}$, the symbol $*$ indicates the binomial thinning operation, so that $\alpha_{it} * y_{i,t-1}$ is the sum of $y_{i,t-1}$ binomial variables with probability α_{it}. Note that the dynamic model in (8.86) is similar but different from the stationary longitudinal model (6.14) used for correlated Poisson responses. Unlike in (6.14), the probability α_{it} in (8.86) is time dependent and also it is considered as a random variable. Further suppose that

1. $y_{i,0} \sim \text{NeBi}(1/c^*, c^* \theta_i)$.
2. $d_{it} \sim \text{NeBi}\{(1-\rho)/c^*, c^* \theta_i\}$, all variables being independent, with $\theta_i = \exp(x_i' \beta)$, where x_i is the $p \times 1$ vector of time independent covariates.
3. $b_j(\alpha_{it})$ denotes the jth binary variable, with probability of success α_{it}; that is, $\text{pr}\{b_j(\alpha_{it}) = 1\} = \alpha_{it} = 1 - \text{pr}\{b_j(\alpha_{it}) = 0\}$.
4. α_{it} follows a beta distribution, namely

$$\alpha_{it} \sim \text{Be}\{\rho/c^*, (1-\rho)/c^*\};$$

that is,

$$g(\alpha_{it}) == \frac{\Gamma(1/c^*)}{\Gamma(\frac{\rho}{c^*})\Gamma(\frac{1-\rho}{c^*})} \alpha_{it}^{(\rho/c^*)-1} (1-\alpha_{it})^{(1-\rho)/c^*-1}, \tag{8.87}$$

for all i and t, with $0 \leq \rho \leq 1$,

Now for given $y_{i,t-1}$ and also α_{it}, it follows that

$$\alpha_{it} * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\alpha_{it}) = z_{it}$$

yields the conditional binomial distribution as

$$z_{it} | y_{i,t-1}, \alpha_{it} \sim \text{Bi}(y_{i,t-1}, \alpha_{it}),$$

independently for all i and t. Next, by using the aforementioned assumption 4, that is, beta distribution (8.87) for α_{it}, one obtains the conditional distribution of z_{it} given $y_{i,t-1}$ as

$$g(z_{it}|y_{i,t-1}) = \frac{y_{i,t-1}!}{z_{it}!(y_{i,t-1}-z_{it})!}\frac{\Gamma(1/c^*)}{\Gamma\{(1-\rho)/c^*\}\Gamma(\rho/c^*)} \times \frac{\Gamma(\rho/c^*+z_{it})\Gamma\{(1-\rho)/c^*+y_{i,t-1}-z_{it}\}}{\Gamma(1/c^*+y_{i,t-1})}, \quad (8.88)$$

for $z_{it} = 0,\ldots,y_{i,t-1}$, which is referred to as the beta-binomial distribution. Hence, the unconditional distribution of z_{it} is given by

$$g(z_{it}) = \sum_{y_{i,t-1}=z_{it}}^{\infty} g_{c^*}(z_{it}|y_{i,t-1})g(y_{i,t-1})$$

$$= \frac{\Gamma(\rho/c^*+z_{it})}{\Gamma(\rho/c^*)\,z_{it}!}\left(\frac{1}{\rho+c^*\theta_i}\right)^{\rho/c^*}\left(\frac{c^*\theta_i}{\rho+c^*\theta_i}\right)^{z_{it}}. \quad (8.89)$$

Because $d_{it}\sim\text{NeBi}\{(1-\rho)/c^*, c^*\theta_i\}$ and $z_{it}\sim\text{NeBi}(\rho/c^*, c^*\theta_i)$, independently for all i and t, it then follows that, marginally, $y_{it}\sim\text{NeBi}(1/c^*, c^*\theta_i)$, independently for all i and t.

To examine the correlation structure, one may show by using the relationship $y_{it} = \alpha_{it} * y_{i,t-1} + d_{it}$, for example, that $y_{it}y_{i,t-2} = y_{i,t-2}(\alpha_{it}\alpha_{i,t-1}y_{i,t-2} + d_{i,t-1}\alpha_{it}) + y_{i,t-2}d_{it}$. As the α_{it}s are independent with $E(\alpha_{it}) = \rho$, for all $t = 1,\ldots,T$, it then follows that $E(y_{it}y_{i,t-2}) = \rho^2(\theta_{i.} + c^*\theta_{i.}^2) + \theta_{i.}^2$, yielding lag 2 correlation $\rho_2 = \rho^2$. By similar calculations, one can show that, for $\ell = 1,\ldots,T-1$, the lag ℓ autocorrelation is given by $\rho_\ell = \rho^\ell$, which is a special case of the general correlation structure

$$C_i(\rho) \equiv C_i(\rho_1,\ldots,\rho_{T-1}) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix}, \quad (8.90)$$

[see also (6.25)] yielding the autocovariance structure

$$\Sigma_i(\beta,c^*,\rho) \equiv \Sigma_i(\beta,c^*,\rho_1,\ldots,\rho_{T-1}) = A_i^{1/2}C_i(\rho_1,\ldots,\rho_{T-1})A_i^{1/2}, \quad (8.91)$$

with $A_i = \text{diag}\{\text{var}(Y_{it})\}$, where by (8.85), $\text{var}(Y_{it}) = \theta_i + c^*\theta_i^2$ for the stationary case. Note that it is clear from (8.90) that the stationary negative binomial counts modelled by (8.86) exhibit the same autocorrelation structure (6.25) as in the stationary Poisson case. This is also same as the autocorrelation structure for the Gaussian data.

8.4.1.1 Estimation of Parameters

8.4.1.1.1 GQL Estimation for β

By using (8.61), for example, one may write the GQL estimating equation for β as

$$\sum_{i=1}^{K} \frac{1'_T \partial \theta_i(\beta)}{\partial \beta} \Sigma_i^{-1}(\beta, c^*, \rho)[y_i - \theta_i(\beta) 1_T] = 0, \tag{8.92}$$

where $\Sigma_i(\beta, c^*, \rho)$ is given by (8.91) and $\theta_i = \exp(x'_i \beta)$ for the stationary negative binomial counts.

8.4.1.1.2 Estimation of c^*

GQL(I) Approach

Similar to (8.44), one may compute the independence assumption based GQL estimate of c^* by solving

$$\sum_{i=1}^{K} \frac{\partial \lambda'_i(\beta, c^*, \rho)}{\partial c^*} \Omega_i^{-1}(\beta, c^*, \rho = 0)[u_i - \lambda_i(\beta, c^*, \rho)] = 0, \tag{8.93}$$

where

$$u_i = [y_{i1}^2, \ldots, y_{it}^2, \ldots, y_{iT}^2, y_{i1} y_{i2}, \ldots, y_{iv} y_{it}, \ldots, y_{T-1} y_{iT}]',$$

with

$$\lambda_{itt} = E[Y_{it}^2] = \theta_i + (c^* + 1)\theta_i^2, \; \lambda_{ivt} = \rho^{|v-t|}[\sigma_{ivv}\sigma_{itt}]^{1/2} + \theta_i^2 = \rho^{|v-t|}[\theta_i + c^* \theta_i^2] + \theta_i^2.$$

When it is pretended that $\rho = 0$, the $\Omega_i(\beta, c^*, \rho = 0)$ in (8.93) takes the diagonal matrix form. That is,

$$\Omega_i(\beta, c^*, \rho = 0) = \text{diag}[\ldots, \text{var}(Y_{it}^2), \ldots, \text{var}[Y_{iv} Y_{it}], \ldots], \tag{8.94}$$

where

$$\begin{aligned} \text{var}[Y_{iv} Y_{it}]_{|\rho=0} &= \sigma_{ivv}\sigma_{itt} + \sigma_{ivv}\theta_i^2 + \sigma_{itt}\theta_i^2 \\ &= [\theta_i + c^* \theta_i^2]^2 + 2[\theta_i + c^* \theta_i^2]\theta_i^2, \end{aligned}$$

and by Exercise 8.2,

$$\begin{aligned} \text{var}(Y_{it}^2) &= E[Y_{it}^4] - \lambda_{itt}^2 \\ &= \theta_i + (6 + 7c^*)\theta_i^2 + (4 + 16c^* + 12c^{*2})\theta_i^3 + (4c^* + 10c^{*2} + 6c^{*3})\theta_i^4. \end{aligned}$$

GQL(N) Approach

In this approach, one solves the estimating equation

$$\sum_{i=1}^{K} \frac{\partial \lambda'_i(\beta, c^*, \rho)}{\partial c^*} \Omega_{iN}^{-1}(\beta, c^*, \rho)[u_i - \lambda_i(\beta, c^*, \rho)] = 0, \tag{8.95}$$

where Ω_{iN} is the normality assumption based covariance matrix of u_i. By (8.66) and (8.68), one first computes the third− and fourth-order raw moments; that is, $\delta_{iut\ell} = E[Y_{iu}Y_{it}Y_{i\ell}]$ and $\phi_{iut\ell m} = E[Y_{iu}Y_{it}Y_{i\ell}Y_{im}]$ as the functions of $\theta_{it} = \theta_i$ and σ_{iut}. Next, these moments are used to compute the appropriate variances and covariances of the elements of the u_i vector. For example,

$$\mathrm{cov}[Y_{iu}^2, Y_{it}Y_{i\ell}] = \phi_{iuut\ell} - \lambda_{iuu}\lambda_{it\ell}.$$

Note that under the present stationary negative binomial longitudinal model

$$\theta_{it} = \exp(x_i'\beta), \text{ and } \sigma_{iut} = \rho^{|t-u|}[\sigma_{iuu}\sigma_{itt}]^{1/2}, \quad (8.96)$$

where, for example, $\sigma_{iuu} = \theta_i + c^*\theta_i^2$.

GQL(CR-I) Approach

Let

$$g_i = [(y_{i1} - \theta_{i1})^2, \ldots, (y_{iT} - \theta_{iT})^2, (y_{i1} - \theta_{i1})(y_{i2} - \theta_{i2}), \\ \ldots, (y_{i,T-1} - \theta_{i,T-1})(y_{iT} - \theta_{iT})].$$

In this approach, one then solves the 'working' GQL estimating equation by

$$\sum_{i=1}^{K} \frac{\partial \sigma_i'(\beta, c^*, \rho)}{\partial c^*} \Omega^{*\,-1}_{i}(\beta, c^*, \rho = 0)[g_i - \sigma_i(\beta, c^*, \rho)] = 0, \quad (8.97)$$

where

$$\sigma_i(\beta, \rho, c^*) = [\sigma_{i11}, \ldots, \sigma_{iTT}, \sigma_{i12}, \ldots, \sigma_{i,T-1,T}]',$$

with σ_{itt} and σ_{iut} as defined in (8.96). Note that as opposed to the conditional serially correlated model [see the construction of $\Omega_i^*(I)$ in (8.71) – (8.73) and Lemma 8.1], the computation of

$$\Omega_i^*(\beta, c^*, \rho = 0) \equiv \Omega_i^*(I)$$

is easier. This is because when $\rho = 0$, under the negative binomial model $\sigma_{its} = 0$, whereas in (8.71), $\sigma_{its} = c\mu_{it}\mu_{is}$. More specifically, under the present model

$$\Omega_i^*(\beta, c^*, \rho = 0) = \mathrm{diag}[\mathrm{var}(Y_{i1} - \theta_{i1})^2, \ldots, \mathrm{var}\{(Y_{i,T-1} - \theta_{i,T-1})(Y_{iT} - \theta_{iT})\}].$$

GQL(CR-N) Approach

Similar to (8.97), in this approach we compute c^* by solving the GQL estimating equation

$$\sum_{i=1}^{K} \frac{\partial \sigma_i'(\beta, c^*, \rho)}{\partial c^*} \Omega^{*\,-1}_{iN}(\beta, c^*, \rho)[g_i - \sigma_i(\beta, c^*, \rho)] = 0, \quad (8.98)$$

which is easier to compute than (8.97). This is because the elements of $\Omega^*_{iN}(\beta, c^*, \rho)$ can be computed by using

$$\begin{aligned}\phi_{iut\ell m} &= E[(Y_{iu}-\theta_{iu})(Y_{it}-\theta_{it})(Y_{i\ell}-\theta_{i\ell})(Y_{im}-\theta_{im})] \\ &= \sigma_{iut}\sigma_{i\ell m}+\sigma_{iu\ell}\sigma_{itm}+\sigma_{ium}\sigma_{it\ell},\end{aligned}$$

[see (8.64)] where $\sigma_{iut} = \theta_{it} + c^*\theta^2_{it}$.

8.4.1.1.3 Moment Estimation of ρ

Note that even though the AR(1) type correlation model is considered in (8.86), a more general correlation structure (8.90) may be fitted in the stationary setup. For the present negative binomial model, the lag ℓ correlation can be estimated by using the moment equation

$$\hat{\rho}_{\ell,M} = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-\ell} y^*_{it}y^*_{i(t+\ell)}/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T} y^{*2}_{it}/KT}, \quad \ell = 1,\ldots,T-1, \tag{8.99}$$

with $y^*_{it} = (y_{it}-\theta_{it})/(\sigma_{itt})^{1/2}$, where

$$\theta_{it} = \exp(x'_{it}\beta), \text{ and } \sigma_{itt} = \theta_{it} + c^*\theta^2_{it},$$

with $\theta_{it} = \theta_i = \exp(x'_i\beta)$ in the stationary case. Note that the formula for ρ estimation by (8.99) is similar to that of the stationary longitudinal count data model, given by (6.27), where $\sigma_{itt} = \theta_{it}$ unlike in (8.99). Further note that this formula in (8.99) is different from (8.54) given under the conditional serially correlation model.

8.4.2 A Data Example: Analyzing Epileptic Count Data by Using Poisson and Negative Binomial Longitudinal Models

Jowaheer and Sutradhar (2002) have revisited the epileptic dataset that was earlier analyzed by Thall and Vail (1990), among others. At each of four successive two-weekly clinic visits, the number of seizures occurring over the previous two weeks was reported by each of a group of 59 epileptics. Summary statistics for this response variable are given in Table 8.15. Variances are much larger than their

Table 8.15 Summary statistics for four two-weekly seizure counts for 59 epileptics.

	Visit			
	1	2	3	4
Sample mean	8.949	8.356	8.441	7.305
Sample variance	220.084	103.785	200.182	93.112

corresponding means, indicating clear overdispersion. Thall and Vail (1990) used the negative binomial model for analysing this overdispersed data, but their approach introduced certain random effects, whereas Jowaheer and Sutradhar (2002) assumed that the four counts for each epileptic follow the general longitudinal autocorrelation structure (8.90). The model considered by Jowaheer and Sutradhar (2002) is easily interpreted with regard to the longitudinal correlations of the repeated responses, the longitudinal correlations can be estimated consistently, and, unlike the approach of Thall and Vail (1990), their approach does not require estimation of multidimensional variance components under a mixed model, which is an extremely difficult problem as it is not easy to check the consistency and efficiency of the variance component estimators. In this section, we provide the results for the epileptic study from Jowaheer and Sutradhar (2002).

We consider five covariates, the intercept (INTC) variable for the ith epileptic at time t, denoted by x_{it1}, the adjuvant treatment (TR) x_{it2} , coded as 0 for placebo and 1 for progabide, baseline seizure rate (BR) x_{it3}, the age of the person in that year (Age) x_{it4}, and the interaction (INTA) x_{it5} between treatment and baseline seizure rate. None of these covariates is time dependent. Thus the mean parameter of the negative binomial distribution for the ith person may be denoted by $\theta_{i\cdot} = \exp(x'_{i\cdot}\beta)$, with $x_{i\cdot} = (x_{it1}, x_{it2}, \ldots, x_{it5})'$ for all $t = 1, \ldots, 4$, which is conformable with the notation used in Section 8.4.1. Here β is the 5×1 vector of regression parameters, and it is of interest to estimate β after taking the longitudinal correlations of the data into account.

Table 8.16 Estimates of regression and overdispersion parameters and their estimated standard errors, as well as estimates of autocorrelations, under both negative binomial and Poisson longitudinal models for the epileptic data.

		Parameters								
Model		INTC	TR	BR	age	INTA	c	ρ_1	ρ_2	ρ_3
NeBi	Estimate	0.458	−0.247	0.027	0.021	0.001	0.514	0.522	0.337	0.203
	SE	0.432	0.152	0.004	0.013	0.005	0.312	–	–	–
Poisson	Estimate	0.486	−0.309	0.021	0.028	0.003	–	0.500	0.353	0.191
	SE	0.021	0.113	0.001	0.006	0.002	–	–	–	–

As far as the GQL estimation of the parameters is concerned, Jowaheer and Sutradhar (2002) have used the GQL estimating equation (8.92) for β, and the GQL(N) estimating equation (8.95) based on

$$u_i = [y_{i1}^2, \ldots, y_{iT}^2]'$$

for c^*, and the moment estimating equation (8.99) to estimate the lag correlations ρ_ℓ. Having chosen starting values of zero for the longitudinal correlations and small positive values for the regression and overdispersion parameters, they estimated the parameters iteratively and obtained the estimates shown in Table 8.16. The results for the Poisson model (i.e., for the negative binomial model with $c^* = 0$) are also shown. The autocorrelation values under both Poisson and negative binomial models are large, indicating high longitudinal correlations. The large value of $\hat{c^*}_{GQL(N)} = 0.514$ confirms that the seizure counts data are highly overdispersed. This overdispersion affects the regression estimates as, except for age and interaction, the regression estimates are generally different under the negative binomial and Poisson models. The negative value for $\hat{\beta}_{2,GQL} = -0.247$ under the negative binomial model indicates that the predicted seizure counts will be less in the treatment group than in the placebo group. The positive estimate for β_4 indicates that, as age increases, it is likely that the individual epileptic will have more seizure counts. Unlike as in Thall and Vail (1990), the interaction between the treatment and the baseline seizure rate does not appear to be significant.

8.4.3 Nonstationary Negative Binomial Correlation Models and Estimation of Parameters

8.4.3.1 First Two Moments Based Negative Binomial Autoregression Model

In a time series setup, Mallick and Sutradhar (2008) have exploited an observation-driven model for nonstationary negative binomial counts and discussed the estimation of the parameters of such a model. In the longitudinal setup, this observation-driven model has the form as in (8.86); that is,

$$y_{it} = \alpha_{it} * y_{i,t-1} + d_{it}, \;\; t = 2, \ldots, T; \; i = 1, \ldots, K, \tag{8.100}$$

but unlike (8.86), we now assume that

$$y_{i1} \sim NeBi(c^{*-1}, c^* \theta_{i1}) \tag{8.101}$$

with $\theta_{i1} = \exp(x'_{i1}\beta)$, where $x_{i1} = (x_{i11}, \cdots, x_{i1p})'$ is the p-dimensional vector of covariates associated with y_{i1}, and

$$d_{it} \sim NeBi(\psi_{it|t-1}, \xi_{it|t-1}), \;\; t = 2, \ldots, T, \tag{8.102}$$

with

$$\psi_{it|t-1} = \frac{(\theta_{it} - \rho\theta_{i,t-1})^2}{c^*(\theta_{it}^2 - \rho\theta_{i,t-1}^2)}, \quad \xi_{it|t-1} = \frac{c^*(\theta_{it}^2 - \rho\theta_{i,t-1}^2)}{(\theta_{it} - \rho\theta_{i,t-1})}, \tag{8.103}$$

where $\theta_{it} = \exp(x'_{it}\beta)$ with $x_{it} = (x_{it1}, \cdots, x_{itp})'$, and $y_{i,t-1}$ and d_{it} are independent. Further note that even though it has been assumed that y_{i1} and d_{it}, $t = 2, \ldots, T$, follow marginally a negative binomial distribution, unlike in the stationary case (Section 8.4.1), it is, however, not easy to derive the marginal distribution of y_{it}, for all $t = 2, \ldots, T$. By the same token, it is difficult to compute the moments of order higher than 2, following the nonstationary dynamic model (8.100). We, however, show below that irrespective of the negative binomial marginal distribution for all y_{it}, these responses satisfying (8.100) have the means and the variances of negative binomial variable.

8.4.3.1.1 Nonstationary Mean−Variance Structure

Lemma 8.2 The repeated count responses satisfying the dynamic model (8.100) and the related assumptions through (8.101) to (8.103), have the means and the variances as given by

$$E(Y_{it}) = \theta_{it} = \exp(x'_{it}\beta), \text{ and } \mathrm{var}(Y_{it}) = \theta_{it} + c^*\theta_{it}^2 = \sigma_{itt}, \tag{8.104}$$

respectively.

Proof of Lemma 8.2: It follows from the model (8.100) that

$$\begin{aligned} E(Y_{it}) &= E_{\alpha_{it}} E_{y_{i,t-1}} E\left[Y_{it} \mid y_{i,t-1}, \alpha_{it}\right] \\ &= E_{\alpha_{it}} E_{y_{i,t-1}} \left[\alpha_{it} Y_{i,t-1} + E(d_{it})\right], \end{aligned} \tag{8.105}$$

and

$$\begin{aligned} \mathrm{var}(Y_{it}) &= E_{\alpha_{it}} \left[\mathrm{var}_{y_{t-1}} E(Y_t \mid \alpha_{it}, y_{i,t-1}) + E_{y_{i,t-1}} \mathrm{var}(Y_{it} \mid \alpha_{it}, y_{i,t-1})\right] \\ &+ \mathrm{var}_{\alpha_{it}} \left[E_{y_{i,t-1}} E(Y_{it} \mid \alpha_{it}, y_{i,t-1})\right], \end{aligned} \tag{8.106}$$

where

$$d_{it} \sim NeBi\left[\psi_{it|t-1}, \xi_{it|t-1}\right]$$

as in (8.102) – (8.103), $y_{i1} \sim NeBi(c^{*-1}, c^*\theta_{i1})$ by (8.101) and α_{it} has beta distribution with probability density function $g(\alpha_{it})$ as defined under the model (8.87).

Based on the above assumptions, one obtains

$$\begin{aligned} E(d_{it}) &= \theta_{it} - \rho\theta_{i,t-1} \\ E(Y_{it} \mid \alpha_{it}, y_{it-1}) &= y_{i,t-1}\alpha_{it} + \theta_{it} - \rho\theta_{i,t-1} \\ \mathrm{var}(Y_{it} \mid \alpha_{it}, y_{i,t-1}) &= y_{i,t-1}\alpha_{it}(1-\alpha_{it}) + \theta_{it} - \rho\theta_{i,t-1} + c^*(\theta_{it}^2 - \rho\theta_{i,t-1}^2) \\ E(\alpha_{it}) &= \rho \\ \mathrm{var}(\alpha_{it}) &= \rho(1-\rho)c^*/(1+c^*). \end{aligned} \tag{8.107}$$

Now, for $t = 2$, by applying (8.107) to (8.105) and (8.106), and by using

$$E[Y_{i1}] = \theta_{i1}, \ \mathrm{var}[Y_{i1}] = \theta_{i1} + c^*\theta_{i1}^2,$$

one obtains

$$E[Y_{i2}] = \theta_{i2}, \ \mathrm{var}[Y_{i2}] = \theta_{i2} + c^*\theta_{i2}^2. \tag{8.108}$$

Consequently, by using

$$E[Y_{i,t-1}] = \theta_{i,t-1},\ \mathrm{var}[Y_{i,t-1}] = \theta_{i,t-1} + c^* \theta_{i,t-1}^2,$$

the repeated applications of (8.107) to (8.105) and (8.106), provides the mean and the variance of y_{it}, for all $t = 1, \ldots, T$, as

$$E[Y_{it}] = \theta_{it},\ \mathrm{var}[Y_{it}] = \theta_{it} + c^* \theta_{it}^2,$$

yielding the lemma.

8.4.3.1.2 Non-stationary Correlation Structure

Let ℓ denote the lag between two responses. Also let $\rho_y(\ell)$ denote the lag ℓ correlation between y_{it} and $y_{i,t-\ell}$ for $\ell = 1, \cdots, t-1$. To derive this ℓth lag correlation, one needs to find the

$$\mathrm{cov}(Y_{it}, Y_{i,t-\ell}) = E(Y_{it} Y_{i,t-\ell}) - \theta_{it} \theta_{i,t-\ell}. \tag{8.109}$$

Note that as α_{it} is a beta variable with $E(\alpha_{it}) = \rho$, and $E(d_{it}) = \theta_{it} - \rho \theta_{i,t-1}$, it then follows from the model (8.100) that

$$\begin{aligned} E(Y_{it} Y_{i,t-\ell}) &= E_{y_{i,t-\ell}} E_{y_{i,t-\ell+1}} \cdots E_{y_{i,t-1}} E\left[Y_{it} Y_{i,t-\ell} \mid y_{i,t-1}, y_{i,t-2}, \cdots, y_{i,t-\ell}\right] \\ &= \rho^{\ell} (\theta_{i,t-\ell} + c^* \theta_{i,t-\ell}^2) + \theta_{it} \theta_{i,t-\ell}, \end{aligned}$$

yielding the covariance in (8.109) by (8.104) as

$$\mathrm{cov}(Y_{it}, Y_{i,t-\ell}) = \rho^{\ell} \sigma_{i,t-\ell,t-\ell}. \tag{8.110}$$

It then follows that the lag ℓ autocorrelation between y_{it} and $y_{i,t-\ell}$ is given by

$$\rho_y(\ell) = \rho^{\ell} \sqrt{\frac{\sigma_{i,t-\ell,t-\ell}}{\sigma_{itt}}}. \tag{8.111}$$

It is clear that this lag ℓ autocorrelation in (8.111) is nonstationary. This is because σ_{itt} in (8.104) is a function of θ_{it} which depends on time-dependent covariate x_{it}. Further note that the correlation structure in (8.111) reduces to the Gaussian AR(1) type autocorrelation structure under the stationary negative binomial model where $x_{it} = x_i$ is considered to be time independent. As far as the range restriction of ρ is concerned, it is clear that for $\psi_{it|t-1}$ and $\xi_{it|t-1}$ in (8.103) to be positive, ρ must satisfy

$$0 < \rho < \min\left\{1, \frac{\theta_{it}}{\theta_{i,t-1}}, \frac{\theta_{it}^2}{\theta_{i,t-1}^2}\right\},\ t = 2, \cdots, T;\ i = 1, \ldots, K. \tag{8.112}$$

8.4.3.2 A Proposed Conditional GQL (CGQL) Estimation Approach

The computation of the higher-order moments under the model (8.100) is complicated, unlike the stationary case (see Section 8.4.1.1), therefore we now use a so-called conditional GQL (CGQL) approach for the estimation of the regression effects β and the overdispersion parameter c^*. As expected, the estimation of the ρ parameter is done by using the unconditional moments. For the purpose of estimation of β and c^*, by using model (8.100), here we provide the conditional moments of y_{it} up to order four conditional on $y_{i,t-1}$, as follows.

Conditional Moments: Conditional on $y_{i,t-1}$, the first− and second-order conditional moments easily follow from the model (8.100). These moments are given by

$$\begin{aligned} E(Y_{it} \mid y_{i,t-1}) &= \theta_{it} + \rho(y_{i,t-1} - \theta_{i,t-1}) = \theta_{it|t-1} \\ E(Y_{it}^2 \mid y_{i,t-1}) &= \delta_1 \rho y_{i,t-1}^2 + \rho y_{i,t-1}(1 - \delta_1 + 2a_{it}) + (a_{it}^2 + a_{it} + c^* b_{it}) \\ &= \lambda_{it|t-1}, \end{aligned} \tag{8.113}$$

where

$$a_{it} = \theta_{it} - \rho\theta_{i,t-1}, \;\; b_{it} = \theta_{it}^2 - \rho\theta_{i,t-1}^2, \;\; \text{and} \;\; \delta_1 = \frac{c^* + \rho}{1 + c^*}.$$

The remaining third− and fourth-order conditional moments are available from Exercise 8.3.

8.4.3.2.1 CGQL Estimation for β

To develop a CGQL estimating equation for β, we first construct a distance vector for the ith individual, namely, $y_i - \mu_{i(c)}$, where $y_i = (y_{i1}, \cdots, y_{it}, \cdots, y_{iT})'$ is the response vector and

$$\begin{aligned} \mu_{i(c)} &= [E(y_{i1}), E(y_{i2} \mid y_{i1}), \cdots, E(y_{it} \mid y_{i,t-1}), \cdots, E(y_{iT} \mid y_{i,T-1})]' \\ &= (\theta_{i1}, \theta_{i2|1}, \cdots, \theta_{it|t-1}, \cdots, \theta_{iT|T-1})', \end{aligned} \tag{8.114}$$

is the conditional expectation of y_i. Note that as y_{i1} is the initial response, θ_{i1} is the marginal mean of y_{i1}. Next, suppose that $\Sigma_{i(c)}$ is the conditional covariance of the response vector. To be specific, the (u,t)th component of this $\Sigma_{i(c)}$ matrix is defined as

$$\sigma_{iut(c)} = \begin{cases} \text{var}(Y_{i1}), & \text{for } u = t = 1 \\ \text{var}(Y_{it} | y_{i,t-1}), & \text{for } u = t = 2, \ldots, T \\ \text{cov}(Y_{iu}, Y_{it} | y_{i,t-1}, \cdots, y_{iu}), & \text{for } u < t \\ \text{cov}(Y_{iu}, Y_{it} | y_{i,u-1}, \cdots, y_{it}), & \text{for } u > t. \end{cases} \tag{8.115}$$

Now by following the GQL estimating equation from Sutradhar (2003, Section 3), the CGQL estimating equation for β may be written as

$$\sum_{i=1}^{K} \frac{\partial \mu'_{i(c)}}{\partial \beta} \Sigma_{i(c)}^{-1} (y_i - \mu_{i(c)}) = 0, \tag{8.116}$$

where $\partial \mu'_{i(c)}/\partial \beta$ is the first-order $p \times T$ derivative matrix of $\mu_{i(c)}$ with respect to β. The formulas for the elements of $\mu_{i(c)}$, $\Sigma_{i(c)}$, and $\partial \mu'_{i(c)}/\partial \beta$ are computed as follows.

(a) Computation of $\mu_{i(c)}$

To compute $\mu_{i(c)}$, one requires $E(Y_{i1}) = \theta_{i1}$ and $E(Y_{it}|y_{i,t-1}) = \theta_{it|t-1}$ for $t = 2,\ldots,T$. The formula for $\theta_{it|t-1}$ is given by (8.113).

(b) Computation of $\Sigma_{i(c)}$

This matrix is computed by using the formulas

$$\begin{aligned} \mathrm{var}(Y_{i1}) &= \theta_{i1} + c^* \theta_{i1}^2 & (8.117)\\ \mathrm{var}(Y_{it}|y_{i,t-1}) &= E(Y_{it}^2|y_{i,t-1}) - \theta_{it|t-1}^2 \text{ for } u = 2,\ldots,T & (8.118)\\ \mathrm{cov}(Y_{iu}, Y_{it}|y_{i,t-1},\cdots,y_{iu}) &= 0 \text{ for } u < t & (8.119)\\ \mathrm{cov}(Y_{iu}, Y_{it}|y_{i,u-1},\cdots,y_{it}) &= 0 \text{ for } u > t & (8.120) \end{aligned}$$

where $E(Y_{it}^2|y_{i,t-1})$ and $\theta_{it|t-1}$ for (8.118) are available from (8.113).

(c) Computation of $\partial \mu'_{i(c)}/\partial \beta$

The derivative matrix is computed by calculating $\partial \theta_{i1}/\partial \beta$ and $\partial \theta_{it|t-1}/\partial \beta$. To be specific, for $k = 1,\ldots,p$,

$$\frac{\partial \theta_{i1}}{\partial \beta_k} = x_{i1k}\theta_{i1} \tag{8.121}$$

$$\frac{\partial \theta_{it|t-1}}{\partial \beta_k} = x_{itk}\theta_{it} - \rho x_{i,t-1,k}\theta_{i,t-1}, \tag{8.122}$$

with $\theta_{it} = \exp(x'_{it}\beta)$.

8.4.3.2.2 CGQL Estimation for c^*

Next we proceed to develop the CGQL estimating equation for the overdispersion parameter c^* as follows. To do this, we first construct a second-order response vector $u_i = (y_{i1}^2,\ldots,y_{it}^2,\ldots,y_{iT}^2)'$ and denote its conditional expectation by

$$\begin{aligned} \lambda_{i(c)} &= \left[E(Y_{i1}^2),\ldots,E(Y_{it}^2|y_{i,t-1}),\ldots,E(Y_{iT}^2|y_{i,T-1})\right]' \\ &= \left[\lambda_{i1},\ldots,\lambda_{it|t-1},\ldots\lambda_{iT|T-1}\right]'. \end{aligned} \tag{8.123}$$

Furthermore, let $\Omega_{i(c)}$ denote the conditional covariance of u_i. That is

$$\omega_{iut(c)} = \begin{cases} \text{var}(Y_{i1}^2), & \text{for } u = t = 1 \\ \text{var}(Y_{it}^2 | y_{i,t-1}), & \text{for } u = t = 2, \cdots, T \\ \text{cov}(Y_{iu}^2, Y_{it}^2 | y_{i,t-1}, \cdots, y_{iu}), & \text{for } u < t \\ \text{cov}(Y_{iu}^2, Y_{it}^2 | y_{i,u-1}, \cdots, y_{it}), & \text{for } u > t. \end{cases} \quad (8.124)$$

The CGQL estimating equation for c^* is similar to that of β in (8.116), which is given by

$$\sum_{i=1}^{K} \frac{\partial \lambda'_{i(c)}}{\partial c^*} \Omega_{i(c)}^{-1} (u_i - \lambda_{i(c)}) = 0, \quad (8.125)$$

where $\partial \lambda'_{i(c)} / \partial c^*$ is the $1 \times T$ first-order derivative matrix of $\lambda_{i(c)}$ with respect to c^*. The formulas for the elements of $\lambda_{i(c)}$, $\Omega_{i(c)}$, and $\partial \lambda'_{i(c)} / \partial c^*$ are computed as follows.

(a) Computation of $\lambda_{i(c)}$

To compute $\lambda_{i(c)}$, one requires $E(Y_{i1}^2) = \theta_{i1} + \theta_{i1}^2 (1 + c^*)$ and $E(Y_{it}^2 | y_{i,t-1})$ for $t = 2, \ldots, T$. The formula for $E(Y_{it}^2 | y_{it-1}) = \lambda_{it|t-1}$ is given by (8.113).

(b) Computation of $\Omega_{i(c)}$

This matrix is computed by using the formulas

$$\text{var}(Y_{i1}^2) = \theta_{i1} + (6 + 7c^*)\theta_{i1}^2 + (4 + 16c^* + 12c^{*2})\theta_{i1}^3 + (4c^* + 10c^{*2} + 6c^{*3})\theta_{i1}^4, \quad (8.126)$$

by Exercise 8.2, and

$$\text{var}(Y_{it}^2 | y_{i,t-1}) = E(Y_{it}^4 | y_{i,t-1}) - \lambda_{it|t-1}^2 \ \text{ for } u = 2, \ldots, T, \quad (8.127)$$

by (8.113) and Exercise 8.3. Furthermore,

$$\text{cov}(Y_{iu}^2, Y_{it}^2 | y_{i,t-1}, \cdots, y_{iu}) = 0 \ \text{ for } u < t \quad (8.128)$$

$$\text{cov}(Y_{iu}^2, Y_{it}^2 | y_{i,u-1}, \cdots, y_{it}) = 0 \ \text{ for } u > t. \quad (8.129)$$

(c) Computation of $\partial \lambda'_{i(c)} / \partial c^*$

The derivative matrix is computed by calculating $\partial \lambda_{i1} / \partial c^*$ and $\partial \lambda_{it|t-1} / \partial c^*$. To be specific,

$$\frac{\partial \lambda_{i1}}{\partial c^*} = \theta_{i1}^2, \tag{8.130}$$

and

$$\frac{\partial \lambda_{it|t-1}}{\partial c^*} = \frac{1}{(1+c^*)^2}\left[\rho(1-\rho)y_{i,t-1}(y_{i,t-1}-1)\right] + \theta_{it}^2 - \rho\theta_{i,t-1}^2. \tag{8.131}$$

8.4.3.2.3 MMs Equation for ρ

Recall from Section 8.4.3.1.2 that the lag 1 corrrelation between y_{it} and $y_{i,t-1}$ is given by $\rho_y(1) = \rho\left[\sigma_{i,t-1,t-1}/\sigma_{i,tt}\right]^{1/2}$, where σ_{itt}, for example, is a function of $\theta_{it} = \exp(x_{it}'\beta)$ and c^*. Thus, to compute a moment estimate for ρ, one may equate the lag 1 sample correlation to its population counterpart, namely, $\rho_y(1)$. To be specific, the moment estimator of ρ, that is, $\hat{\rho}_M$ has the formula given by

$$\hat{\rho}_M = \frac{\sum_{i=1}^K \sum_{t=2}^T \tilde{y}_{it}\tilde{y}_{i,t-1}}{\sum_{i=1}^K \sum_{t=1}^T \tilde{y}_{it}^2} \frac{KT}{\sum_{i=1}^K \sum_{t=2}^T \left[\hat{\sigma}_{i,t-1,t-1}/\hat{\sigma}_{i,tt}\right]^{1/2}}, \tag{8.132}$$

where

$$\tilde{y}_{it} = \frac{y_{it} - \hat{\theta}_{it}}{\sqrt{\hat{\sigma}_{itt}}}.$$

Exercises

8.1. (Section 8.3.2.1) [Construction of $\Omega_i^*(I)$]
The formulae for $\text{var}[(Y_{it}-\mu_{it})^2]$ and $\text{cov}[(Y_{it}-\mu_{it})^2, (Y_{it}-\mu_{it})(Y_{is}-\mu_{is})]$ were given in Lemma 8.1. For

$$r_{iut} = \left[\frac{\mu_{iu}}{\mu_{it}}\right]^{\frac{1}{2}}$$

and

$$\rho_{|t-s|} = \rho^{|t-s|},$$

by similar calculations as in Lemma 8.1, show that the other elements of the $\Omega_i^*(I)$ matrix have a general formula given by

$$\begin{aligned}
&\text{cov}[(Y_{it}-m_{it})(Y_{iw}-m_{iw}), (Y_{ir}-m_{ir})(Y_{is}-m_{is})] \\
&= \exp(\sigma_\gamma^2)\left[r_{itw}r_{its}\rho_{|t-w|}\rho_{|r-s|} + r_{itr}r_{iws}\rho_{|t-r|}\rho_{|w-s|}\right. \\
&\quad \left. + r_{iwr}r_{its}\rho_{|w-r|}\rho_{|t-s|}\right]\exp\{\frac{1}{2}(x_{it}+x_{iw}+x_{ir}+x_{is})'\beta\} \\
&\quad + \left\{\exp(3\sigma_\gamma^2) - 2\exp(\sigma_\gamma^2) + 1\right\}\left[r_{itw}\rho_{|t-w|}\exp\{(x_{it}/2 + x_{iw}/2 + x_{ir} + x_{is})'\beta\}\right.
\end{aligned}$$

$$+r_{irs}\rho_{|r-s|}\exp\{(x_{it}+x_{iw}+x_{ir}/2+x_{is}/2)'\beta\}$$

$$+r_{itr}\rho_{|t-r|}\exp\{(x_{it}/2+x_{iw}+x_{ir}/2+x_{is})'\beta\}$$

$$+r_{iws}\rho_{|w-s|}\exp\{(x_{it}+x_{iw}/2+x_{ir}+x_{is}/2)'\beta\}$$

$$+r_{iwr}\rho_{|w-r|}\exp\{(x_{it}+x_{iw}/2+x_{ir}/2+x_{is})'\beta\}$$

$$+r_{its}\rho_{|t-s|}\exp\{(x_{it}/2+x_{iw}+x_{ir}+x_{is}/2)'\beta\}]$$

$$+\{\exp(6\sigma^2)-4\exp(3\sigma^2)+6\exp(\sigma^2)-3\}\exp\{(x_{it}+x_{iw}+x_{ir}+x_{is})'\beta\}$$

$$-\sigma_{itw}\sigma_{irs}, \tag{8.133}$$

When the formula in (2.16) is evaluated at $\rho_0 = 1$ and $\rho_{|t-w|} = 0$ for $t \neq w$, it provides the formulas for all elements of the $\Omega_i^*(I)$ matrix except the formulae for the variance and covariance provided in (8.75) and (8.76), respectively.

8.2. (Section 8.4.1.1) [Higher-order marginal moments for negative binomial distribution]
Show that for the negative binomial distribution of y_{it}, the moment generating function (mgf) is given by

$$M_{y_{it}}(s) = \{1 + c\theta_{it} - c\theta_{it}\exp(s)\}^{-1/c},$$

where s is a real parameter. Also by using the mgf, verify that

$$\begin{aligned}
\text{var}(Y_{it}) &= \theta_{it} + c\theta_{it}^2,\\
\text{cov}(Y_{it}, Y_{it}^2) &= \theta_{it}\{1 + (2+3c)\theta_{it} + 2c(1+c)\theta_{it}^2\},\\
\text{var}(Y_{it}^2) &= \theta_{it} + (6+7c)\theta_{it}^2 + (4+16c+12c^2)\theta_{it}^3 + (4c+10c^2+6c^3)\theta_{it}^4.
\end{aligned}$$

8.3. (Section 8.4.3.2.1) [Third− and fourth-order conditional moments]
Let

$$\delta_2 = \frac{2c^* + \rho}{1 + 2c^*}, \text{ and } \delta_3 = \frac{3c^* + \rho}{1 + 3c^*}.$$

Also, for convenience, suppress the subscript i, and use a_t and b_t, for $a_{it} = \theta_{it} - \rho\theta_{i,t-1}$ and $b_{it} = \theta_{it}^2 - \rho\theta_{i,t-1}^2$, respectively. Now by using the model (8.100), show that conditional on y_{t-1}, (i suppressed) the third− and the fourth-order moments of y_t (i suppressed) are given by

$$\begin{aligned}
E(Y_t^3 \mid y_{t-1}) &= \delta_1\delta_2\rho y_{t-1}^3 + \delta_1\rho y_{t-1}^2(3 + 3a_t - 3\delta_2)\\
&\quad + \rho y_{t-1}\left[1 + 6a_t + 3a_t^2 + 3c^* b_t - \delta_1(3 + 3a_t - 2\delta_2)\right]
\end{aligned}$$

$$+\frac{1}{a_t}(a_t^4+3a_t^3+a_t^2+3c^*a_tb_t+3c^*a_t^2b_t+2c^{*2}b_t^2), \quad (8.134)$$

and

$$\begin{aligned} E(Y_t^4 \mid y_{t-1}) = {} & \delta_1\rho y_{t-1}^2(6a_t^2+18a_t+6c^*b_t+7) \\ & +\rho y_{t-1}\left[\delta_1\delta_2\delta_3(y_{t-1}-1)(y_{t-1}-2)(y_{t-1}-3)\right. \\ & +\delta_1\delta_2(y_{t-1}-1)(y_{t-1}-2)(4a_t+6) \\ & +4a_t(1-3\delta_1)+(1-7\delta_1)+6(1-\delta_1)(a_t^2+a_t+c^*b_t) \\ & \left.+\frac{4}{a_t}\left(a_t^4+3a_t^3+a_t^2+3c^*a_tb_t+3c^*a_t^2b_t+2c^{*2}b_t^2\right)\right] \\ & +\frac{1}{a_t^2}\left(a_t^6+6a_t^5+7a_t^4+a_t^3+10c^*a_t^3b_t+14c^*a_t^4b_t\right. \\ & +7c^*a_t^2b_t+12c^{*2}a_tb_t^2+3c^{*2}a_t^2b_t^2+8c^{*2}a_t^3b_t^2 \\ & \left.+6c^{*3}b_t^3\right), \end{aligned} \quad (8.135)$$

respectively.

References

1. Blundell, R., Griffith, R., & Windmeijer, F. (1995). Individual effects and dynamics in count data. *Discussion paper* 95-03. Department of Economics, University College London.
2. Hansen, L. -P. (1982). Large sample properties of generalized method of moment estimators. *Econometrica,* 50, 1029 – 1054.
3. Hausman, J. A., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R and D relationship. *Econometrica,* 52, 908 – 938.
4. Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.,* 93, 720 – 729.
5. Jowaheer, V. & Sutradhar, B. C. (2002). Analysing longitudinal count data with overdispersion. *Biometrika,* 89, 389 – 399.
6. Jowaheer, V. & Sutradhar, B. C. (2009). GMM versus GQL inferences for panel count data. *Statist. Probab. Lett.,* 79, 1928 – 1934.
7. Lewis, P. A. W. (1980). Simple models for positive-valued and discrete-valued time series with ARMA correlation structure. In *Multivariate Analysis*, ed. P. R. Krishnaiah, pp. 151 – 166. Amsterdam: North Holland.
8. Mallick, T. S. & Sutradhar, B. C. (2008). GQL versus conditional GQL inferences for non-sttaionary time series of counts with overdispersion. it J. Time Series Anal., 29, 402 – 420.
9. McKenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric marginal distributions. *Adv. App. Prob.,* 18, 679 – 705.
10. Montalvo, J. G. (1997). GMM Estimation of count-panel-data models with fixed effects and predetermined instruments. *J. Busi. and Econo. Stats.,* 15, 82 – 89.

11. Prentice, R.L. & Zhao, L.P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics,* 47, 825 – 839.
12. Sutradhar, B. C. (2003). An Overview on Regression Models for Discrete Longitudinal Responses. *Statist. Sci.,* 18, 377 – 393.
13. Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhya,* 66, 261 – 289.
14. Sutradhar, B. C., Bari, W. (2007). On generalized quasilikelihood inference in longitudinal mixed model for count data. *Sankhya,* 69, 671 – 699.
15. Sutradhar, B. C. & Jowaheer, V. (2003). On familial longitudinal Poisson mixed models with gamma random effects. *J. Multivariate Anal.,* 87, 398 – 412.
16. Thall, P. F. & Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics,* 46, 657 – 671.
17. Wooldridge, J. (1999). Distribution-free estimation of some non-linear panel data models. *J. Econometrics,* 90, 77 – 97.

Appendix

U.S. Patent Data For 168 Industries : Table 8A
Corresponding Covariates Including R&D Expenditures : Table 8B

Table 8A. U.S. patent data from 168 industries from 1974 to 1979.

Patent Awarded					
1974	1975	1976	1977	1978	1979
2	3	2	1	1	1
0	0	0	1	0	0
8	0	0	0	1	0
5	2	3	0	2	3
1	0	1	5	2	0
1	0	0	0	0	0
12	4	3	5	5	5
0	2	1	2	4	2
4	3	3	0	2	2
2	1	0	0	1	2
0	2	1	1	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1	4	3	4	2	6
4	3	4	2	1	1
1	0	3	0	2	0
1	1	0	0	0	0
6	4	4	1	4	3
1	0	1	0	0	0
2	5	0	1	3	0
5	5	8	6	2	5
0	0	2	0	0	0
0	0	0	0	0	0
1	0	2	0	0	1
1	0	2	0	1	2
7	5	5	10	2	4
0	1	1	2	2	1
0	0	3	2	1	0
5	5	7	7	5	9
4	2	5	2	3	5
2	4	4	2	13	2
3	4	3	11	4	5
1	1	5	2	3	1
1	0	0	1	0	0
1	5	1	2	4	0
0	3	0	7	7	9
6	6	4	4	3	3
8	3	2	1	2	8
0	1	2	2	1	0

Table 8A Cont'd

1974	1975	1976	1977	1978	1979
0	0	2	0	2	1
2	3	2	2	0	3
3	5	6	6	8	4
8	4	3	4	13	4
0	0	1	1	0	0
0	0	2	1	1	0
0	3	5	2	1	1
9	10	6	11	13	7
1	2	1	1	2	6
0	0	0	0	0	0
0	2	0	2	3	0
9	3	6	13	5	7
1	4	5	6	2	4
2	6	7	1	1	1
9	13	4	2	2	1
1	0	2	0	0	0
0	0	0	0	0	0
3	4	2	2	4	1
8	2	4	3	0	1
1	0	2	1	5	1
3	0	2	0	2	4
10	1	3	3	5	3
1	1	1	0	0	0
2	0	4	3	3	2
7	2	3	0	1	0
3	1	2	2	2	1
1	2	3	3	4	3
1	1	4	5	0	1
1	1	0	0	0	0
1	0	0	1	1	0
2	1	2	0	1	0
1	1	1	5	5	8
2	1	3	1	0	0
0	0	1	0	0	0
3	4	0	0	4	3
2	2	4	5	2	4

Table 8A Cont'd

1974	1975	1976	1977	1978	1979
2	1	3	2	1	0
0	0	1	1	0	0
15	13	6	4	9	6
2	2	3	0	2	0
0	0	1	0	4	0
5	2	1	0	0	1
7	3	3	5	8	13
1	2	1	0	2	0
0	0	0	1	0	0
0	0	0	0	0	1
4	10	6	6	2	1
6	6	1	1	3	2
16	14	8	6	12	12
7	6	6	1	0	1
9	5	10	4	5	5
1	0	0	0	1	0
0	0	0	0	0	0
5	6	2	6	7	5
5	6	4	5	2	5
7	5	3	5	3	4
0	0	0	0	0	0
2	0	2	2	2	4
0	0	0	1	0	0
1	0	0	0	0	0
0	0	0	0	0	1
0	1	0	1	0	1
1	0	0	0	0	0
0	2	1	0	0	0
10	7	4	6	2	7
0	1	0	1	0	0
6	1	5	1	9	5
3	5	1	4	3	2
0	3	1	0	0	0
0	0	0	0	0	0
0	1	0	0	0	1
0	0	1	0	1	0
4	2	2	0	1	6
5	4	4	3	4	2
0	3	2	4	0	3
3	1	0	3	1	0

Table 8A Cont'd

1974	1975	1976	1977	1978	1979
4	3	0	0	0	0
2	3	3	0	0	1
4	5	1	0	1	0
2	2	2	1	1	4
4	2	4	8	5	5
3	0	1	5	1	9
2	1	1	1	2	3
0	0	0	0	1	0
0	1	0	1	0	0
11	6	7	7	9	9
2	5	3	2	2	1
3	1	3	4	7	5
4	1	2	2	0	0
5	2	8	2	6	5
1	2	2	0	0	2
0	0	0	0	0	0
1	0	0	0	0	0
0	6	5	3	1	0
4	1	4	2	3	1
4	1	0	5	7	1
2	4	0	1	1	3
0	6	0	4	2	1
0	1	2	5	1	3
0	0	0	0	0	0
0	0	0	0	0	0
4	12	6	10	6	5
0	1	1	1	0	0
3	0	3	1	3	2
15	12	11	13	10	5
11	7	4	6	6	2
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
7	1	0	4	9	5
0	0	2	0	0	2
5	4	2	0	1	1
1	2	2	0	0	0
4	3	3	8	1	2
4	5	3	0	0	1
4	2	6	6	6	2
9	3	0	2	4	5
1	0	4	0	0	0
3	1	1	1	1	2
2	5	3	4	6	2
0	0	0	0	1	0
2	1	1	1	1	0
8	0	1	1	4	1
0	0	0	0	0	0
3	4	5	3	8	5
3	2	1	1	2	3
3	0	0	1	0	0
10	10	20	6	12	12
16	11	11	12	13	14

Table 8B. Covariates corresponding to patent counts in Table 8A.

Industry Type	Book Value	R & D Expenditures 1971	1972	1973	1974	1975	1976	1977	1978	1979
1	1.975	−0.216	0.084	−0.151	−0.685	−1.485	−1.195	−0.610	−0.581	−0.609
0	0.684	0.485	0.588	0.488	0.537	0.434	0.338	0.366	0.439	0.425
0	2.064	−0.889	−0.315	−0.218	−0.362	−1.298	−1.675	−2.150	−1.325	−2.834
1	2.790	0.224	0.357	0.371	0.041	−0.318	−0.190	−0.199	0.079	0.194
0	0.678	−0.593	−2.733	−1.714	−1.361	−1.800	−1.732	−1.560	−1.544	−1.560
0	3.819	−0.244	−0.435	−0.708	−0.758	−0.721	−1.001	−0.483	−0.358	−0.422
1	3.644	0.878	1.002	0.907	0.910	0.958	1.053	1.191	1.287	1.513
0	2.455	−1.325	−1.076	−0.937	−0.883	−0.936	−0.902	−0.905	−0.910	−1.000
1	2.493	0.617	0.536	0.432	0.370	0.079	−0.157	−0.260	−0.081	−0.188
0	2.955	0.450	0.405	0.274	-0.092	0.355	0.214	0.296	0.508	0.450
0	0.406	−0.133	−0.182	−0.211	−0.616	−0.691	−1.543	−1.272	−1.379	−1.450
0	3.674	−0.936	−0.870	−0.811	−0.740	−0.740	−0.622	−0.676	−0.604	−0.489
0	1.434	−1.019	−1.044	−1.162	−1.227	−1.229	−1.171	−0.798	−0.684	−0.548
0	3.002	0.236	0.278	0.297	0.399	0.352	0.502	0.593	0.661	0.838
0	3.534	−0.348	−0.232	−0.128	−0.056	0.020	0.152	0.186	0.172	0.229
0	1.507	−0.822	−1.204	−2.007	−2.708	−2.253	−2.153	−2.775	−2.697	−2.425
0	3.767	0.503	0.650	0.612	0.587	0.495	0.484	0.532	0.522	0.686
0	3.826	1.967	1.963	1.547	1.844	2.011	2.309	2.289	2.307	2.262
0	3.022	−0.466	−0.163	−0.075	−0.528	−0.668	−0.627	−0.637	−0.757	−0.047
1	3.383	2.037	2.113	2.076	1.818	1.954	1.910	1.886	1.935	1.962
1	−0.249	−0.218	−0.167	−0.140	−0.078	0.176	0.543	0.599	0.566	0.462
0	0.832	−0.240	−0.828	−0.267	−0.046	−0.013	0.097	−0.005	0.113	0.139
0	1.590	−1.447	−1.050	−0.537	−0.308	−0.497	−0.672	−0.459	−0.492	−0.636
0	1.048	−1.438	−1.402	−1.269	−1.391	−1.538	−1.585	−1.442	−1.311	−1
0	3.611	0.148	0.087	0.312	0.392	−0.068	−0.398	−0.486	−1.062	−1.683
0	3.840	1.463	1.463	1.193	0.515	0.387	0.601	0.625	0.736	0.736
0	1.959	−0.241	−0.136	−0.032	0.105	−0.261	−0.408	−0.434	0.275	0.121
1	2.370	0.633	0.896	0.987	0.329	0.402	0.401	0.420	0.217	0.033
1	3.211	0.522	0.655	0.989	0.945	0.892	0.939	0.942	0.855	0.854
0	2.932	0.279	0.427	0.461	0.559	0.781	0.945	0.951	0.997	0.935
1	1.759	−0.762	−0.693	−0.755	−0.473	−0.271	−0.945	−0.905	−0.278	−0
1	2.905	1.275	1.075	1.359	1.582	1.300	1.761	1.868	2.141	2.093

Table 8B Cont'd

Industry Type	Book Value	1971	1972	1973	1974	1975	1976	1977	1978	1979
1	3.489	1.323	1.398	1.573	1.623	1.604	1.719	1.909	1.980	2.067
0	2.835	0.276	70.255	0.330	0.322	0.188	0.247	0.298	0.159	0.123
1	2.711	0.878	0.833	0.978	0.869	0.815	1.217	1.554	1.562	1.678
0	2.027	−0.991	−1.109	−1.733	−1.631	−1.779	−1.731	−1.626	−1.565	−1.552
1	3.506	1.320	1.441	1.555	1.551	1.565	1.577	1.611	1.649	1.688
1	2.209	0.141	0.391	0.382	0.198	0.416	−0.050	0.186	0.357	0.338
0	0.698	−1.276	−1.938	−1.523	−1.391	−1.468	−1.546	−1.269	−1.136	−1.220
1	2.334	0.633	0.759	0.730	−0.512	−0.370	0.424	−0.476	−1.073	−0.714
0	1.628	−0.118	0.307	0.467	0.238	0.303	−0.072	−0.050	0.151	−0.062
1	0.609	0.773	1.030	0.748	0.051	−0.243	−0.424	0.082	0.270	0.457
0	2.402	−0.689	−0.796	−0.684	−0.790	−0.829	−0.669	−0.843	−0.861	−0.960
1	3.439	−1.088	−1.019	−0.930	−0.491	−0.680	−0.507	−0.580	−0.623	−0.533
1	1.686	0.527	0.781	0.112	−0.399	−0.840	−0.943	−0.889	−1.657	−2.704
1	3.198	0.738	0.596	0.431	0.542	0.515	0.573	0.567	0.684	0
1	3.758	2.605	2.560	2.366	2.371	2.517	2.507	2.687	2.766	2.778
0	2.464	0.086	0.451	0.813	0.562	0.696	0.679	0.671	0.360	0.218
0	1.442	0.450	0.833	0.678	1.087	1.128	1.264	2.038	2.146	1.742
1	0.863	−2.258	−1.766	−1.276	−1.073	−1.081	−0.649	−0.544	−0.299	−0.346
1	1.268	−0.061	0.244	0.261	0.069	0.114	0.181	0.310	0.732	1.130
0	3.197	1.380	0.840	1.085	1.087	0.553	0.128	0.174	0.104	0.010
0	1.816	−1.345	−1.366	−1.647	−0.768	−1.579	−0.879	−2.366	−2.441	−2.020
0	2.279	0.410	0.452	0.680	0.614	0.492	0.744	0.669	0.633	0.738
1	−1.633	−3.531	−2.976	−2.459	−2.364	−2.563	−2.872	−2.942	−2.918	−3.006
1	1.928	−0.008	0.028	0.070	−0.071	0.083	0.129	0.237	0.208	0.246
1	2.828	0.450	0.313	0.438	0.009	−0.586	−0.381	−0.310	−0.279	0.395
0	2.315	−0.312	−0.942	−1.043	−0.416	−0.493	−0.515	−1.733	−1.387	−1.647
1	1.053	−0.897	−0.875	−0.821	−0.881	−0.794	−0.691	−0.143	−0.171	−0.252
0	2.097	0.227	0.331	0.287	0.329	0.200	0.298	0.453	0.504	0.467
1	2.802	0.738	0.531	0.340	0.879	0.207	0.451	0.544	0.602	0.967

Table 8B Cont'd

Industry Type	Book Value	1971	1972	1973	1974	1975	1976	1977	1978	1979
0	0.806	−2.067	−1.884	−1.846	−1.844	−2.486	−3.304	−2.680	−2.168	−1.340
1	2.690	−0.433	−0.046	−0.192	−0.030	0.028	−0.052	−0.019	0.124	0.055
0	3.123	−0.078	−0.281	−0.104	−0.035	−0.147	−0.026	−0.067	−0.129	−0.119
1	2.427	0.518	0.634	0.763	0.913	0.782	0.821	0.960	1.076	1.137
0	3.359	1.102	0.745	0.981	0.962	0.951	1.028	1.056	1.015	0.906
0	0.407	−1.159	−1.523	−1.421	−1.416	−1.445	−0.826	−0.525	−0.494	−0.447
1	1.458	−0.128	−0.183	−0.021	−0.176	0.056	0.287	0.389	0.322	0.325
1	1.718	−0.135	−0.035	−0.047	0.066	−0.039	0.016	0.454	0.566	0.633
0	3.169	−0.528	−0.409	−1.293	−0.036	−0.461	0.012	0.262	0.189	−0.071
0	1.031	−0.869	−0.870	−0.467	−0.492	−0.579	−0.595	−0.442	−0.232	0.154
1	1.196	0.178	0.292	0.427	0.671	0.815	1.049	0.788	0.420	0.414
1	1.749	−0.494	−0.949	−1.279	−1.432	−1.457	−1.730	−2.403	−2.697	−3.577
0	1.907	−0.889	−0.952	−0.988	−0.588	−0.762	−0.681	−0.404	−0.468	−0.389
1	2.963	0.511	0.448	0.526	0.641	0.545	0.677	0.553	0.668	0.673
1	1.204	0.410	0.696	0.978	1.096	1.267	1.087	1.140	1.245	1.429
0	0.253	−2.432	−3.352	−3.674	−3.153	−3.548	−3.849	−3.479	−3.112	−3.092
1	2.375	0.630	0.468	0.594	0.593	0.400	0.546	0.722	0.827	1.176
0	0.718	−1.362	−1.221	−2.158	−1.742	−1.880	−1.305	−1.287	−0.972	−0.759
0	0.351	0.009	0.137	0.399	0.588	−0.780	−0.529	−0.228	−0.090	0.601
0	2.517	0.450	−0.564	0.498	0.284	0.014	−0.141	−0.461	−0.226	−0.550
0	3.861	−0.734	−0.701	−0.656	−0.897	−0.916	−1.502	−1.644	−1.580	−0.082
0	−0.741	−1.447	−1.478	−1.353	−1.272	−1.216	−1.242	−1.190	−1.077	−0.076
0	2.718	0.094	−0.233	0.479	−0.308	−0.706	−0.734	−1.669	−1.792	−1.647
0	2.172	−2.820	−3.124	−3.141	−2.906	−2.693	−2.581	−2.044	−1.684	−1.659
1	1.475	−0.055	−0.050	0.454	0.700	0.744	0.543	0.662	0.810	0.829
1	0.829	−0.758	−0.468	−0.505	−0.450	−0.503	−0.492	0.097	0.461	0.535
1	3.969	1.579	1.604	1.662	1.690	1.688	1.875	2.025	2.023	2.020
1	2.021	0.410	0.461	0.527	0.727	1.012	1.269	1.398	1.500	1.588
0	3.621	0.961	1.132	1.569	1.722	1.652	1.733	1.715	1.740	1.776
0	0.321	−1.407	−2.033	−1.211	−2.174	−1.678	−0.766	−1.137	−0.508	−0.270

Table 8B Cont'd

Industry Type	Book Value	1971	1972	1973	1974	1975	1976	1977	1978	1979
0	2.519	−2.026	−2.323	−2.693	−2.500	−2.524	−2.471	−2.282	−2.093	−1.745
0	3.494	0.099	0.330	−0.568	−0.064	−0.417	−0.304	−0.410	−0.521	−0.662
0	1.241	−0.176	0.109	−0.051	−0.103	−0.035	0.126	0.294	0.181	0.206
0	2.950	0.576	0.594	0.580	0.538	0.391	0.434	0.721	0.808	0.878
0	3.297	0.278	0.292	0.325	0.327	0.198	0.135	0.160	0.107	−0
1	1.786	0.009	0.588	0.480	0.518	0.231	0.570	0.449	0.824	0.796
0	2.859	−1.425	−1.461	−1.502	−1.459	−1.524	−1.592	−1.602	−1.583	−1.516
0	1.507	−3.013	−2.526	−2.216	−2.303	−2.615	−2.716	−2.603	−2.590	−2.121
0	3.001	−1.616	−1.715	0.631	0.536	0.666	0.790	0.931	0.800	0.794
1	0.552	−1.139	−0.794	−0.581	−0.582	−0.405	−0.351	−0.377	−0.458	−0.240
0	2.833	−1.079	−1.291	−1.409	−1.073	−1.788	−1.800	−1.764	−1.369	−1.522
1	1.456	0.190	−0.152	−0.170	−0.675	−0.582	−0.550	−0.474	−0.879	−0.911
1	3.054	1.208	1.194	1.324	1.287	1.051	1.072	1.084	1.098	1.430
0	1.959	−0.771	−1.386	−1.266	−1.766	−1.747	−1.667	−1.702	−1.701	−1.745
1	1.896	−0.244	−0.543	−0.234	−0.108	−0.076	−0.087	0.344	0.606	0.484
0	2.229	−0.907	−0.919	−0.775	−0.683	−0.631	−0.682	−0.512	−0.377	−0.215
1	3.182	−1.005	−0.203	−0.308	−0.439	−0.831	0.028	0.153	0.189	0.201
0	3.203	−1.893	−2.137	−2.141	−2.002	−2.524	−2.173	−2.680	−2.669	−2.574
0	2.109	−2.506	−2.590	−2.110	−2.430	−2.371	−2.034	−2.426	−2.557	−2.262
0	2.833	−0.104	−0.511	−0.629	−1.245	−1.506	−0.766	−0.732	−0.810	−0.490
1	0.271	−0.957	0.170	0.567	0.647	1.073	0.852	1.211	1.591	1.555
0	2.900	−0.957	−1.204	−1.448	−0.262	−0.081	0.222	0.264	0.289	0
1	2.521	0.450	0.149	0.079	0.019	−0.168	−0.109	−0.115	−0.159	−0.193
0	2.650	−0.406	−0.361	−0.374	−0.310	−0.108	−0.028	0.094	0.135	0.099
1	2.384	0.512	−0.229	−0.138	−0.123	−0.176	−0.193	0.113	0.143	−0.320
1	2.197	0.045	0.123	0.326	0.451	0.461	0.402	0.460	0.516	0.441
1	−0.121	−2.519	−2.235	−1.360	−1.204	−1.063	−1.007	−1.599	−0.956	−0.225
0	0.796	−2.408	−1.214	−1.421	−1.855	−2.053	−2.311	−2.380	−2.304	−2.301
1	3.114	1.065	1.057	1.075	1.113	0.866	0.334	0.637	0.658	0.653
0	3.804	1.038	1.024	0.894	0.834	0.962	1.003	1.104	1.242	1.290
0	1.885	0.214	0.259	0.043	0.330	0.073	0.282	0.492	0.592	0.683
0	2.788	−0.596	−0.417	−0.199	−0.274	−0.244	−0.343	−0.279	−0.080	0.196
1	−0.053	−1.037	−1.041	−1.307	−1.338	−1.422	−1.473	−0.947	−0.850	−0.905
0	3.472	0.700	0.668	0.702	0.730	0.494	0.506	0.606	0.580	0.673
0	2.839	−1.113	−0.732	−1.816	−1.416	−1.372	−1.740	−2.054	−1.851	−2.064
0	2.769	0.112	0.125	0.150	0.298	0.012	0.047	0.100	0.280	0.485
0	3.601	0.687	0.649	0.430	0.340	0.571	0.347	0.222	−0.057	−0.0101
0	3.659	1.390	1.533	1.440	1.436	1.429	1.537	0.222	1.615	1.606
0	3.517	0.673	0.723	0.655	0.613	0.621	0.628	0.709	0.636	0.789
0	1.316	−2.971	−1.394	−1.554	−1.699	−1.583	−1.679	−1.539	−1.506	−2.438
1	2.532	−0.371	−0.473	−0.556	−0.730	−0.739	−0.817	−0.780	−0.687	−0.637
1	1.529	−0.881	−0.260	−0.090	−0.345	−0.065	0.205	0.404	1.004	1.164

Table 8B Cont'd

Industry Type	Book Value	1971	1972	1973	1974	1975	1976	1977	1978	1979
1	1.642	−1.681	−0.728	−0.510	−0.910	−1.047	−1.222	−1.309	−1.142	−1.008
0	1.042	−0.648	−0.916	−0.897	−0.688	−0.653	−0.662	−0.730	−0.722	−0.749
1	−1.079	−1.506	−1.465	−1.401	−1.426	−1.090	−1.162	−1.312	−1.313	−1.135
0	1.389	−0.996	−0.504	−0.504	−0.269	−0.171	−0.104	−0.538	−0.346	−0.543
0	2.738	−0.113	0.044	0.187	0.316	0.159	0.384	0.468	0.542	0
0	3.625	−2.420	−2.303	−2.365	−2.460	−2.553	−2.515	−2.388	−2.286	−2.307
0	2.856	−1.341	−2.283	−2.208	−1.971	−0.995	−0.936	−0.894	−0.828	−0.812
0	3.697	0.193	0.421	0.638	0.607	0.056	0.277	0.160	0.123	0.188
0	2.076	−0.876	−1.191	−1.293	−2.167	−1.917	−1.964	−1.764	−1.464	−0.833
1	3.106	0.461	0.544	0.566	0.623	0.607	0.674	0.801	0.899	0.911
1	3.955	2.038	0.878	1.678	1.803	1.378	1.397	1.524	1.535	1.678
1	2.266	−0.243	−0.301	−0.213	−0.224	−0.365	−0.640	−0.784	−0.655	−0.660
0	2.099	−1.739	−1.448	−2.065	−2.683	−2.841	−2.953	−2.786	−2.991	−2.983
1	1.741	1.304	0.226	0.093	0.482	0.546	0.516	0.941	1.173	1.367
0	0.557	−2.162	−0.904	−1.712	−2.802	−3.246	−1.943	−1.883	−3.283	−3.388
1	3.692	1.605	1.665	1.519	0.518	0.510	0.723	1.260	1.576	1.646
1	−1.770	−1.727	−1.988	−1.432	−1.474	−1.475	−1.333	−1.616	−1.523	−1.114
0	2.688	0.227	0.339	0.194	−0.343	0.151	0.281	0.315	0.371	0.349
0	2.021	0.237	0.179	0.175	0.032	0.025	0.240	0.247	0.414	0.555
0	2.864	−0.245	−0.047	0.070	0.076	0.122	−0.100	0.105	0.248	0.319
0	2.698	0.381	0.588	0.580	0.485	0.337	0.434	0.539	0.463	0.488
1	2.762	0.700	0.702	0.772	0.753	0.741	0.754	0.882	1.074	1.274
0	3.386	0.376	0.592	0.579	0.576	0.229	0.258	0.269	0.283	0.382
1	3.238	0.920	0.742	−0.755	−0.365	−1.321	−1.659	−2.156	−2.524	−3.388
0	2.190	−1.564	−1.124	−0.708	−0.219	−0.251	−0.046	0.028	0.017	−0.010
1	1.972	0.082	−0.223	−0.364	−0.145	−0.074	−0.089	−0.087	0.720	0.068
1	1.806	−0.089	−0.087	0.079	0.153	−0.058	0.076	0.156	0.255	0.593
0	2.245	−0.310	−0.188	−0.056	−0.062	−0.054	−0.028	0.244	0.161	0.076
1	2.443	−0.006	0.155	0.118	0.106	0.108	0.084	0.068	0.040	0.195
0	0.659	−3.307	−3.058	−3.058	−2.460	−2.984	−2.051	−2.498	−3.112	−3.275
0	2.718	0.576	0.742	0.825	0.818	0.608	0.713	0.628	0.640	0.633
1	2.972	−0.256	−0.105	−0.062	0.158	0.207	0.222	0.315	0.503	0.524
1	2.178	−1.159	−0.892	−0.978	−0.768	−1.289	−1.384	−1.009	−1.013	−1.748
0	3.754	0.873	0.835	0.782	0.739	0.666	0.936	0.916	0.898	0.750
1	3.373	1.837	1.885	2.024	2.101	2.119	2.236	2.401	2.489	2.480

Chapter 9
Longitudinal Mixed Models for Binary Data

Recall that various stationary and nonstationary correlated binary fixed models were discussed in Chapter 7. In this chapter, we consider a generalization of some of these fixed models to the mixed model setup by assuming that the repeated binary responses of an individual may also be influenced by the individual's random effect. Thus, this generalization will be similar to that for the repeated count data subject to the influence of the individual's random effect that we have discussed in Chapter 8. Note that in this chapter, we concentrate mainly on the nonstationary models, stationary models being the special cases.

In Section 9.1, we discuss a binary longitudinal mixed model as a generalization of the linear dynamic nonstationary AR(1) model used in Section 7.4.1. The basic properties as well as the estimation of the parameters of the mixed model are also given. In Section 9.2, we provide a generalization of the nonlinear binary dynamic logit (BDL) model discussed in Section 7.7.2, to the mixed model setup. This generalized model is referred to as the binary dynamic mixed logit (BDML) model, the BDL model being alternatively referred to as the binary dynamic fixed logit (BDFL) model. The so-called IMM (improved method of moments) and GQL (generalized quasi-likelihood) estimation approaches are discussed in detail for the estimation of the parameters, namely the regression effects and dynamic dependence parameter as well as the variance of the random effect, of the BDML model. We revisit the SLID data analyzed by fitting the BDFL model in Section 7.5, and reanalyze it now by fitting the BDML model. In the same section, we also include the likelihood estimation and compare its performance with the GQL approach. In Section 9.3, we consider a binary dynamic mixed probit (BDMP) model as an alternative to the BDML model and use the GQL estimation approach for the desired misspecification inferences.

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_9,

9.1 A Conditional Serially Correlated Model

Let $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ be the T repeated binary responses collected from the ith $(i = 1, \ldots, K)$ individual, $x_{it} = (x_{it1}, \ldots, x_{itj}, \ldots, x_{itp})'$ be the p-dimensional covariate vector associated with the response y_{it}, and $\beta = (\beta_1, \ldots, \beta_j, \ldots, \beta_p)'$ denote the regression effects of x_{it} on y_{it}. Because the repeated responses are likely to be correlated, in Chapter 7, more specifically in Section 7.4, they were modelled based on a class of nonstationary autocorrelation structures, namely AR(1), MA(1), and EQC (equicorrelations). In this section, we, for example, consider the nonstationary AR(1) model only. The other models may be treated similarly. However, in addition to the stochastic time effect, we now assume that the repeated binary responses of an individual are also influenced by the individual random effect. Consequently, conditional on the random effect $\gamma_i \overset{i.i.d.}{\sim} N(0, \sigma_\gamma^2)$, the repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are assumed to follow the AR(1) correlation model

$$
\begin{aligned}
Pr[Y_{i1} = 1|\gamma_i] &= \pi_{i1}^* \\
Pr[Y_{it} = 1|\gamma_i, y_{i,t-1}] &= \pi_{it}^* + \rho(y_{i,t-1} - \pi_{i,t-1}^*), \text{ for } t = 2, \ldots, T,
\end{aligned}
\tag{9.1}
$$

where $\pi_{it}^* = \exp(x_{it}'\beta + \gamma_i)/[1 + \exp(x_{it}'\beta + \gamma_i)]$, for all $t = 1, \ldots, T$.

9.1.1 Basic Properties of the Model

Conditional on the random effects γ_i, the linear dynamic probability model (9.1) yields the conditional means and the variances as

$$
\begin{aligned}
E(Y_{it}|\gamma_i) &= \pi_{it}^* \\
\text{var}(Y_{it}|\gamma_i) &= \sigma_{itt}^* = \pi_{it}^*(1 - \pi_{it}^*),
\end{aligned}
\tag{9.2}
$$

for $t = 1, \ldots, T$. Next, for $u < t$, by using the model relationship (9.1), similar to (7.72), one may compute the conditional covariance between y_{iu} and y_{it} as

$$
\text{cov}[(Y_{iu}, Y_{it})|\gamma_i] = \rho^{t-u}\sigma_{iuu}^*. \tag{9.3}
$$

The unconditional means, variances and covariances may be obtained as in the following lemma.

Lemma 9.1. By using

$$
\gamma_i^* = \frac{\gamma_i}{\sigma_\gamma} \overset{\text{iid}}{\sim} N(0, 1)
$$

so that $\pi_{it}^*(\gamma_i^*) = \exp(x_{it}'\beta + \sigma_\gamma\gamma_i^*)/[1 + \exp(x_{it}'\beta + \sigma_\gamma\gamma_i^*)]$, and the conditional moments from (9.2) and (9.3), one obtains the unconditional means, variances, and covariances as

$$E[Y_{it}] = \pi_{it}(\beta, \sigma_\gamma^2) = \int \pi_{it}^*(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* \tag{9.4}$$

$$\text{var}[Y_{it}] = \sigma_{itt}(\beta, \sigma_\gamma^2) = \pi_{it}(\beta, \sigma_\gamma^2)(1 - \pi_{it}(\beta, \sigma_\gamma^2)) \tag{9.5}$$

$$\begin{aligned}
\text{cov}[Y_{iu}, Y_{it}] &= \sigma_{iut}(\beta, \sigma_\gamma^2, \rho) \\
&= \rho^{t-u}\left[\pi_{iu} - \int \pi^{*2}_{iu}(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^*\right] \\
&\quad + \left[\int \pi_{iu}^*(\gamma_i^*)\pi_{it}^*(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* - \pi_{iu}\pi_{it}\right] \\
&= \rho^{t-u}[\pi_{iu} - \pi_{iuu}] + [\pi_{iut} - \pi_{iu}\pi_{it}],
\end{aligned} \tag{9.6}$$

with $g_N(\gamma_i^*|1)$ as the standard normal density, yielding the pairwise familial correlations as

$$\text{corr}[Y_{ij}, Y_{ik}] = \frac{\rho^{t-u}[\pi_{iu} - \pi_{iuu}] + [\pi_{iut} - \pi_{iu}\pi_{it}]}{[\pi_{iu}(\beta, \sigma_\gamma^2)(1 - \pi_{iu}(\beta, \sigma_\gamma^2))\pi_{it}(\beta, \sigma_\gamma^2)(1 - \pi_{it}(\beta, \sigma_\gamma^2))]^{1/2}}. \tag{9.7}$$

Proof: In the manner similar to that of Lemma 5.1, one obtains the unconditional mean, variance, and the covariance in (9.4) – (9.6) by using the following formulas.

$$\begin{aligned}
E[Y_{it}] &= E[Y_{it}^2] = E_{\gamma_i^*} E[Y_{it}|\gamma_i^*] \\
\text{var}[Y_{it}] &= E_{\gamma_i^*}[\text{var}\{Y_{it}|\gamma_i^*\}] + \text{var}_{\gamma_i^*}[E\{Y_{it}|\gamma_i^*\}] \\
\text{cov}[Y_{iu}, Y_{it}] &= E_{\gamma_i^*}\text{cov}[\{Y_{iu}, Y_{it}\}|\gamma_i^*] + \text{cov}_{\gamma_i^*}[E(Y_{iu}|\gamma_i^*), E(Y_{it}|\gamma_i^*)],
\end{aligned} \tag{9.8}$$

where by (9.3), the conditional covariance, that is, $\text{cov}[\{Y_{iu}, Y_{it}\}|\gamma_i^*]$ is a function of the longitudinal correlation index parameter ρ.

Note that similar to that in Chapter 5, π_{iuu} and π_{iut} in (9.6) may be computed by using either a simulation or binomial approximation. To be specific, in the simulation technique, for a large N such as $N = 1000$, π_{iuu} in (9.6), for example, may be computed as

$$\pi_{iuu}^{(s)}(\beta, \sigma_\gamma^2) = \frac{1}{N}\sum_{w=1}^{N}[\pi_{iu}^{*2}(\gamma_{iw}^*)], \tag{9.9}$$

[see also (5.20)] where γ_{iw}^* is a sequence of standard normal values for $w = 1, \ldots, N$. Alternatively, one may approximate the desired normal integral by a binomial approximation and compute π_{iuu} as

$$\pi_{iuu}^{(b)}(\beta, \sigma_\gamma^2) = \sum_{v_i=0}^{V} \pi_{iu}^{*2}(v_i)] \binom{V}{v_i} (1/2)^{v_i}(1/2)^{V-v_i}, \tag{9.10}$$

[see also (5.24)] where for a known reasonably big V such as $V = 5$,

$$v_i \sim \text{binomial}(V, 1/2),$$

and hence it has a relation to γ_i^* as

$$\gamma_i^* = \frac{v_i - V(1/2)}{V(1/2)(1/2)}.$$

Further note that unlike in Chapter 5, π_{iut} in (9.6) is not the same as $\lambda_{iut} = E[Y_{iu}Y_{it}]$. This is because when y_{iu} and y_{it} are correlated,

$$\lambda_{iut} = E[Y_{iu}Y_{it}] \neq E_{\gamma_i^*}[\pi_{iu}^* \pi_{it}^*],$$

when $\rho \neq 0$.

9.1.2 Parameter Estimation

9.1.2.1 GQL Estimation of the Regression Effects β

For $y_i = [y_{i1}, \ldots, y_{it}, \ldots, y_{iT}]'$, let

$$E[Y_i] = \pi_i(\beta, \sigma_\gamma^2) = [\pi_{i1}(\beta, \sigma_\gamma^2), \ldots, \pi_{it}(\beta, \sigma_\gamma^2), \ldots, \pi_{iT}(\beta, \sigma_\gamma^2)]',$$

with

$$\begin{aligned}\pi_{it}(\beta, \sigma_\gamma^2) &= \int \frac{\exp(x_{it}'\beta + \sigma_\gamma \gamma_i^*)}{1 + \exp(x_{it}'\beta + \sigma_\gamma \gamma_i^*)} g_N(\gamma_i^*|1) d\gamma_i^* \\ &= \int \pi_{it}^*(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* \\ &= \pi_{it}^{(b)}(\beta, \sigma_\gamma^2), \text{ (say).} \end{aligned} \tag{9.11}$$

Next, let $\Sigma_i(\beta, \sigma_\gamma^2, \rho)$ denote the covariance matrix of y_i. To be specific,

$$\Sigma_i(\beta, \sigma_\gamma^2, \rho) = (\sigma_{iut}), \tag{9.12}$$

where $\text{var}(Y_{it}) = \sigma_{itt} \equiv \sigma_{itt}(\beta, \sigma_\gamma^2)$ and $\text{cov}(Y_{iu}, Y_{it}) = \sigma_{iut} \equiv \sigma_{iut}(\beta, \sigma_\gamma^2, \rho)$ for $u \neq t$, with σ_{itt} and σ_{iut} defined as in (9.5) and (9.6), respectively. Now following (8.18) [see also Sutradhar (2004)], one may solve the estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \pi_i'(\beta, \sigma_\gamma^2)}{\partial \beta} \Sigma_i^{-1}(\beta, \sigma_\gamma^2, \rho)[y_i - \pi_i(\beta, \sigma_\gamma^2)] = 0, \tag{9.13}$$

to obtain the GQL estimate of β. In (9.13), the first derivative vector may be computed simply by using the formula for the derivative of $\pi_{it}(\beta,\ \sigma_\gamma^2)$ with respect to β. This formula for the derivative is given by

$$\begin{aligned}\frac{\partial \pi_{it}(\beta,\ \sigma_\gamma^2)}{\partial \beta} &= \int \frac{\partial \pi^*_{it}(\gamma^*_i)}{\partial \beta} g_N(\gamma^*_i|1)d\gamma^*_i \\ &= x_{it}\int [\pi^*_{it}(\gamma^*_i)\{1-\pi^*_{it}(\gamma^*_i)\} g_N(\gamma^*_i|1)d\gamma^*_i \\ &= x_{it}[pi^{(b)}_{it}(\beta,\sigma^2_\gamma) - \pi^{(b)}_{itt}(\beta,\ \sigma^2)].\end{aligned}$$

Note that for given σ_γ^2 and ρ, the GQL estimate obtained from (9.13) is consistent for β. This is because, as $E(Y_i) = \pi_i(\beta,\sigma_\gamma^2)$, the estimating equation (9.13) is unbiased. Furthermore, because the GQL estimating equation (9.13) is constructed by using the covariance matrix $\Sigma_i(\beta,\sigma_\gamma^2,\rho)$ as a weight matrix, it follows that the GQL estimate of β obtained from (9.13) would be highly efficient as compared to other competitors such as the method of moments based estimate.

9.1.2.2 GQL Estimation of the Random Effects Variance σ_γ^2

For the estimation of σ_γ^2, the GQL approach exploits the squared and the pairwise product of the observations. Let

$$u_i = (y_{i1}^2,\ldots,y_{iT}^2, y_{i1}y_{i2},\ldots,y_{it}y_{i,t+1},\ldots,y_{i,T-1}y_{iT})'$$

with its expectation

$$\begin{aligned}\lambda_i(\beta,\sigma_\gamma^2,\rho) &= E[U_i] \\ &= (\lambda_{i11},\ldots,\lambda_{itt},\ldots,\lambda_{iTT},\lambda_{i12},\ldots,\lambda_{iut},\ldots,\lambda_{i,T-1,T})'. \quad (9.14)\end{aligned}$$

Because y_{it}^2 and y_{it} are the same in the binary case, to compute $\lambda_i(\beta,\sigma_\gamma^2,\rho)$ one uses

$$\begin{aligned}\lambda_{itt} &\equiv \lambda_{itt}(\beta,\sigma_\gamma^2) = \pi_{it} \\ \lambda_{iut} &\equiv \lambda_{iut}(\beta,\sigma_\gamma^2,\rho) = E(Y_{iu}Y_{it}) \\ &= \rho^{t-u}[\pi_{iu} - \pi_{iuu}] + \pi_{iut}, \quad (9.15)\end{aligned}$$

by (9.6), for all $u < t$. By using the QL principle similar to that of (8.18), one may now write the GQL estimating equation for σ_γ^2 as

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta,\sigma_\gamma^2,\rho)}{\partial \sigma_\gamma^2} \Omega_i^{-1}(\beta,\sigma_\gamma^2,\rho)[u_i - \lambda_i(\beta,\sigma_\gamma^2,\rho)] = 0, \quad (9.16)$$

[Sutradhar and Jowaheer (2003)] where Ω_i is the covariance matrix of u_i. Note that it is, however, extremely cumbersome to compute Ω_i in general under the autoregression model (9.1). As a remedy, we consider two approximations, namely replacing the Ω_i matrix by a 'working' independence assumption based fourth-order moments matrix $\Omega_i(I)$, or replacing the Ω_i matrix by a 'working' normality based weight matrix Ω_{iN}.

9.1.2.2.1 GQL(I) Estimation of σ_γ^2

Under the 'working' independence based approximation, one solves the 'working' GQL(I) estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta, \sigma_\gamma^2, \rho)}{\partial \sigma_\gamma^2} \Omega_{iw}^{-1}(\beta, \sigma_\gamma^2, \rho = 0)[u_i - \lambda_i(\beta, \sigma_\gamma^2, \rho)] = 0, \tag{9.17}$$

for the estimation of σ_γ^2.

Computation for $\Omega_{iw}(\beta, \sigma_\gamma^2, \rho = 0)$

The computation of this matrix requires the computations of the third− and fourth-order product moments. By using the binomial approximation [see (9.10)], for example, the third order moments may be computed as

$$\begin{aligned}
\delta_{iut\ell} &= E[Y_{iu}Y_{it}Y_{i\ell}] \\
&= E_{\gamma_i^*}[E(Y_{iu}|\gamma_i^*)E(Y_{it}|\gamma_i^*)E(Y_{i\ell}|\gamma_i^*)] \\
&= E_{\gamma_i^*}[\pi_{iu}^*\pi_{it}^*\pi_{i\ell}^*] \\
&= \sum_{v_i=0}^{V} [\pi_{iu}^*(v_i)\pi_{it}^*(v_i)\pi_{i\ell}^*(v_i)] \binom{V}{v_i} (1/2)^{v_i}(1/2)^{V-v_i} \\
&= \pi_{iut\ell}^{(b)}(\beta, \sigma_\gamma^2).
\end{aligned} \tag{9.18}$$

Similarly, the fourth-order moments under the assumption that $\rho = 0$ may be computed as

$$\begin{aligned}
\phi_{iut\ell m} &= E[Y_{iu}Y_{it}Y_{i\ell}Y_{im}] \\
&= E_{\gamma_i^*}[E(Y_{iu}|\gamma_i^*)E(Y_{it}|\gamma_i^*)E(Y_{i\ell}|\gamma_i^*)E(Y_{im}|\gamma_i^*)] \\
&= E_{\gamma_i^*}[\pi_{iu}^*\pi_{it}^*\pi_{i\ell}^*\pi_{im}^*] \\
&= \sum_{v_i=0}^{V} [\pi_{iu}^*(v_i)\pi_{it}^*(v_i)\pi_{i\ell}^*(v_i)\pi_{im}^*(v_i)] \binom{V}{v_i} (1/2)^{v_i}(1/2)^{V-v_i} \\
&= \pi_{iut\ell m}^{(b)}(\beta, \sigma_\gamma^2).
\end{aligned} \tag{9.19}$$

9.1.2.2.2 GQL(N) Estimation of σ_γ^2

Under the 'working' normality based approximation, one solves the 'working' GQL(N) estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta, \sigma_\gamma^2, \rho)}{\partial \sigma_\gamma^2} \Omega_{iN}^{-1}(\beta, \sigma_\gamma^2, \rho)[u_i - \lambda_i(\beta, \sigma_\gamma^2, \rho)] = 0, \tag{9.20}$$

for the estimation of σ_γ^2.

Computation for $\Omega_{iN}(\beta, \sigma_\gamma^2, \rho)$

Similar to Section 8.3.1.2, the third− and fourth-order moments for the correlated binary variables, under the normality assumption, are computed as follows. Note that under normality

$$E[(Y_{iu} - \pi_{iu})(Y_{it} - \pi_{it})(Y_{i\ell} - \pi_{i\ell})] = 0, \tag{9.21}$$

yielding the third-order raw moments as

$$\delta_{iut\ell} = E[Y_{iu} Y_{it} Y_{i\ell}] = \sigma_{iut}\pi_{i\ell} + \sigma_{iu\ell}\pi_{it} + \sigma_{it\ell}\pi_{iu} - 2\pi_{iu}\pi_{it}\pi_{i\ell}, \tag{9.22}$$

where by (9.4), (9.6), and (9.10), one writes

$$\begin{aligned} \pi_{it} &\equiv \pi_{it}^{(b)} \\ \sigma_{iut} &= \sigma_{iut}^{(b)} = \rho^{t-u}\left[\pi_{iu}^{(b)} - \pi_{iuu}^{(b)}\right] + \left[\pi_{iut}^{(b)} - \pi_{iu}^{(b)}\pi_{it}^{(b)}\right]. \end{aligned} \tag{9.23}$$

Similarly, because under normality

$$E[(Y_{iu} - \pi_{iu})(Y_{it} - \pi_{it})(Y_{i\ell} - \pi_{i\ell})(Y_{im} - \pi_{im})] = \sigma_{iut}\sigma_{i\ell m} + \sigma_{iu\ell}\sigma_{itm} + \sigma_{ium}\sigma_{it\ell}, \tag{9.24}$$

one obtains the fourth-order raw product moments as

$$\begin{aligned} E[Y_{iu}Y_{it}Y_{i\ell}Y_{im}] &= \phi_{iut\ell m} \\ &= \sigma_{iut}\sigma_{i\ell m} + \sigma_{iu\ell}\sigma_{itm} + \sigma_{ium}\sigma_{it\ell} \\ &\quad + \delta_{iut\ell}\pi_{im} + \delta_{iutm}\pi_{i\ell} + \delta_{iu\ell m}\pi_{it} + \delta_{it\ell m}\pi_{iu} \\ &\quad - \sigma_{iut}\pi_{i\ell}\pi_{im} - \sigma_{iu\ell}\pi_{it}\pi_{im} - \sigma_{ium}\pi_{it}\pi_{i\ell} - \sigma_{it\ell}\pi_{iu}\pi_{im} \\ &\quad - \sigma_{itm}\pi_{iu}\pi_{i\ell} - \sigma_{i\ell m}\pi_{iu}\pi_{it} + 3\pi_{iu}\pi_{it}\pi_{i\ell}\pi_{im}, \end{aligned} \tag{9.25}$$

where π_{it} and σ_{iut}, for example, are given by (9.23), and $\delta_{iut\ell}$ is given by (9.22).

9.1.2.3 Estimation of ρ Under the GQL Approach

Note that the regression effect β may be estimated by using the GQL estimating equation (9.13), and the variance of the random effects σ_γ^2 may be estimated by using either the GQL(I) estimating equation in (9.17) or the GQL(N) equation in (9.20), provided ρ is known. But in practice ρ is rarely known. For given β and σ_γ^2, the correlation or probability parameter (ρ) may be consistently estimated by solving a suitable moment estimating equation that may be developed by equating the population covariance of the data given in (9.6) with its sample counterpart. Note that as ρ is a correlation parameter under the autoregressive order 1 setup, similar to the Gaussian set up, it would be sufficient to exploit the lag 1 autocovariance only to estimate this parameter. More specifically, as by (9.5)

$$\mathrm{var}[Y_{it}] = \sigma_{itt}(\beta, \sigma_\gamma^2) = \pi_{it}(\beta, \sigma_\gamma^2)(1 - \pi_{it}(\beta, \sigma_\gamma^2)),$$

and by (9.6)

$$\mathrm{cov}[Y_{it}, Y_{i,t+1}] = \sigma_{it,t+1}(\beta, \sigma_\gamma^2, \rho) = \rho\,[\pi_{it} - \pi_{itt}] + [\pi_{it,t+1} - \pi_{it}\pi_{i,t+1}]\,,$$

in the manner similar to that of the Poisson mixed model case [see eqn. (8.54)], ρ may be estimated consistently by

$$\hat{\rho} = \frac{a_1 - b_1}{g_1}, \tag{9.26}$$

where a_1 is the observed lag 1 correlation defined as

$$a_1 = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-1} y_{it}^* y_{i(t+1)}^* / K(T-1)}{\sum_{i=1}^{K} \sum_{t=1}^{T} y_{it}^{*2} / KT},$$

with $y_{it}^* = (y_{it} - \mu_{it})/(\sigma_{itt})^{1/2}$, where $\sigma_{itt} = \pi_{it}[1 - \pi_{it}]$. In (9.26),

$$g_1 = \frac{1}{K(T-1)} \sum_{i=1}^{K} \sum_{t=1}^{T-1} \left[\frac{\pi_{it} - \pi_{itt}}{(\sigma_{itt}\sigma_{i,t+1,t+1})^{1/2}} \right],$$

and

$$b_1 = \frac{1}{K(T-1)} \sum_{i=1}^{K} \sum_{t=1}^{T-1} \left[\frac{\pi_{it,t+1} - \pi_{it}\pi_{i,t+1}}{(\sigma_{itt}\sigma_{i,t+1,t+1})^{\frac{1}{2}}} \right].$$

9.2 Binary Dynamic Mixed Logit (BDML) Model

As opposed to the linear binary dynamic mixed model considered in Section 9.1, in this section, mainly following Sutradhar, Rao, and Pandit (2008), we consider a nonlinear binary dynamic mixed model, given by

$$Pr(y_{it} = 1|\gamma_i) = \begin{cases} \frac{\exp(x_{i1}'\beta + \gamma_i)}{1 + \exp(x_{i1}'\beta + \gamma_i)} = p_{i10}^*, & \text{for } i = 1, \ldots, K;\ t = 1 \\ \frac{\exp(x_{it}'\beta + \theta y_{i,t-1} + \gamma_i)}{1 + \exp(x_{it}'\beta + \theta y_{i,t-1} + \gamma_i)} = p_{ity_{i,t-1}}^*, & \text{for } i = 1, \ldots, K;\ t = 2, \ldots, T \end{cases}$$

$$= F_{it}, \text{ say,} \tag{9.27}$$

where β is the regression effects of fixed covariates, θ is the dynamic dependence parameter, and $\gamma_i \overset{i.i.d.}{\sim} N(0, \sigma_\gamma^2)$ is the latent random effect of the ith $(i = 1, \ldots, K)$ individual. Note that for convenience, we use $\gamma_i^* = \gamma_i/\sigma_\gamma$. Further note that the binary mixed model in (9.27) is a direct generalization of the BDFL model considered in Section 7.7.2 [see also Sutradhar and Farrell (2007)]. Also, this model in (9.27) is known as the binary panel data model in the econometrics literature. Many authors such as Heckman (1981), Manski (1987), and Honore and Kyriazidou (2000) studied this model, for a distribution-free random effects case.

For the inferences for β and θ, Honore and Kyriazidou (2000, p. 844), for example, attempted to estimate these parameters by exploiting the first differences of the responses $y_{i1} - y_{i0}$, $y_{i2} - y_{i1}, \ldots$, which are approximately independent of γ_i. For example, for a special case with $T = 4$, they suggest to estimate β and θ by maximizing an approximate weighted log-likelihood function

$$\log\tilde{L} = \sum_{i=1}^{I} I_\delta\{y_{i2} + y_{i3} = 1\} I_\delta\{x_{i3} - x_{i4} = 0\}$$
$$\times\, ln\left(\frac{\exp((x_{i2} - x_{i3})\beta + \theta(y_{i1} - y_{i4}))^{y_{i2}}}{1 + \exp((x_{i2} - x_{i3})\beta + \theta(y_{i1} - y_{i4}))}\right), \qquad (9.28)$$

which seems to be very restrictive as, in longitudinal setup, it is unlikely that x_{i3} will be the same as x_{i4} to yield the indicator function value $I_\delta\{x_{i3} - x_{i4} = 0\} = 1$. As a remedy to this problem due to nonstationarity, Honore and Kyriazidou (2000, eqn. 6, p. 845) further suggest to replace the indicator function $I_\delta\{x_{i3} - x_{i4} = 0\} = 1$ by a kernel density function $\kappa\{(x_{i3} - x_{i4})/b_K\}$, where b_K is the bandwidth that shrinks as K increases. This replacement, however, appears to be quite artificial in order to avoid the technical difficulty produced by the method. In fact for larger T, the estimation problem will be much more difficult. Thus, even if one is interested in the estimation of β and θ, this semiparametric approach of Honore and Kyriazidou (2000) appears to be impractical.

Note that in practice, unlike Honore and Kyriazidou (2000), one may be interested to have an idea about the dispersion (σ_γ^2) of the random effects, as this parameter affects both the mean and the variance of the binary responses. It is clear that obtaining the likelihood estimators of β, θ, and σ_γ^2 in (9.27) requires the maximization of the exact likelihood function

$$L(\beta,\, \theta,\, \sigma_\gamma) = \int_{\gamma_1=-\infty}^{\infty} \cdots \int_{\gamma_K=-\infty}^{\infty} \prod_{i=1}^{K} \{F_{i1}(x'_{i1}\beta + \sigma_\gamma \gamma_i^*\}^{y_{i1}}$$
$$\times \{1 - F_{i1}(x'_{i1}\beta + \sigma_\gamma \gamma_i^*)\}^{1-y_{i1}}$$
$$\times \left[\prod_{t=2}^{T} \{F_{it}(x'_{it}\beta + \theta y_{i,t-1} + \sigma_\gamma \gamma_i^*)\}^{y_{it}}\right.$$

$$\times \left\{1 - F_{it}(x'_{it}\beta + \theta y_{i,t-1} + \sigma_\gamma \gamma_i^*)\right\}^{1-y_{it}}\Big]$$

$$\times\phi(\gamma_1^*)\ldots\phi(\gamma_K^*)\, d\gamma_1^* \ldots d\gamma_K^*, \tag{9.29}$$

which appears to be manageable but complicated. In (9.29), $\phi(\gamma_i^*)$ is the standard normal density, and F_{it} is the conditional probability given by (9.27).

As opposed to the aforementioned complex weighted likelihood and exact likelihood estimation approaches, in the following section we discuss a generalized method of moments (GMM) (referred to as the IMM in Section 5.2.2) that produces consistent estimates for the parameters $\beta,\ \theta,$and σ^2_γ involved in the BDML model. In Section 9.2.2, we discuss the GQL estimation for the same parameters and demonstrate that the GQL approach is much more efficient as compared to the GMM estimation approach. Note that these GMM and GQL approaches were developed by Sutradhar, Rao, and Pandit (2008) for the estimation of the parameters in this BDML model.

9.2.1 GMM/IMM Estimation

As pointed out earlier, the GMM approach due to Hansen (1982) is a popular estimation approach in the econometrics literature. For example, see the articles in the 'Twentieth Anniversary GMM Issue' of the *Journal of Business and Economic Statistics*. Let $\alpha = (\beta', \theta, \sigma^2_\gamma)'$ be the $(p+2)$-dimensional vector of the parameters of the dynamic mixed model (9.27). The construction of the GMM estimating equations for the components of α requires the formulas for their unbiased estimating functions. These unbiased functions are given in the next section.

9.2.1.1 Construction of the Unbiased Moment Functions

Note that in the binary panel data model (9.27), β is the regression parameter vector and θ is the scalar dynamic dependence parameter, whereas σ^2_γ is the variance component of the random effects. Let

$$\psi_i(y_i, \beta,\ \theta,\ \sigma^2_\gamma) = [\psi'_{1i}, \psi_{2i}, \psi_{3i}]' \tag{9.30}$$

be a vector of three unbiased moment functions corresponding to three parameters $\beta,\ \theta,$ and σ^2_γ. Now to construct the first and third components of this vector, one may refer to the construction of the moment functions under the binary mixed model discussed in Chapter 5 (see Section 5.2.2). This is because, for the case when $\theta = 0$, the present BDML model reduces to the BDFL model which has been exploited extensively in the statistics literature to analyze binary data in the generalized linear mixed model (GLMM) setup. For example, for the estimation of β and σ^2_γ, Jiang (1998) [see also Jiang and Zhang (2001) and Sutradhar (2004)] has exploited the sufficient statistics under the conditional GLMM set up and constructed the basic distance functions as

$$
\begin{aligned}
\psi_{1i} &= \sum_{t=1}^{T} x_{it}[y_{it} - \pi_{it}], \text{ and} \\
\psi_{3i} &= \sum_{t=1}^{T}[y_{it}^2 - \lambda_{itt}] + \sum_{u<t}^{T}[y_{iu}y_{it} - \lambda_{iut}] \\
&= \sum_{t=1}^{T}[y_{it} - \pi_{it}] + \sum_{u<t}^{T}[y_{iu}y_{it} - \lambda_{iut}], \qquad (9.31)
\end{aligned}
$$

respectively, where $\pi_{it} = \lambda_{itt} = E[Y_{it}]$ and $\lambda_{iut} = E[Y_{iu}Y_{it}]$; their formulas are given as follows.

9.2.1.1.1 Formula for π_{it}

It follows from the BDML model (9.27) that conditional on γ_i^*, the means of the repeated binary responses are given by

$$
\pi_{it}^*(\gamma_i^*) = E[Y_{it}|\gamma_i^*] = \begin{cases} \frac{\exp(x_{i1}'\beta+\sigma_\gamma\gamma_i^*)}{1+\exp(x_{i1}'\beta+\sigma_\gamma\gamma_i^*)}, & \text{for } i = 1,\ldots,I;\ t = 1 \\ p_{it0}^* + \pi_{i,t-1}^*(p_{it1}^* - p_{it0}^*), & \text{for } i = 1,\ldots,I;\ t = 2,\ldots,T \end{cases} \qquad (9.32)
$$

[see also (7.145)] where

$$
p_{it1}^* = \frac{\exp(x_{it}'\beta + \theta + \sigma_\gamma\gamma_i^*)}{[1+\exp(x_{it}'\beta + \theta + \sigma_\gamma\gamma_i^*)]} \text{ and } p_{it0}^* = \frac{\exp(x_{it}'\beta + \sigma_\gamma\gamma_i^*)}{[1+\exp(x_{it}'\beta + \sigma_\gamma\gamma_i^*)]}.
$$

Subsequently, one obtains the unconditional means as

$$
\begin{aligned}
\pi_{it} &= E(Y_{it}) = Pr(y_{it} = 1) \\
&= M^{-1}\sum_{w=1}^{M}\pi_{it}^*(\gamma_{iw}^*) \\
&= M^{-1}\sum_{w=1}^{M}[p_{it0}^* + \pi_{i,t-1}^*(p_{it1}^* - p_{it0}^*)]_{|\gamma_i^*=\gamma_{iw}^*} \qquad (9.33)
\end{aligned}
$$

[Jiang (1998); Sutradhar (2004)] where γ_{iw}^* is the wth $(w = 1,\ldots,M)$ realized value of γ_i^* generated from the standard normal distribution. Here M is a sufficiently large number, such as $M = 5000$. By (9.32), the $p_{it1,w}^*$ involved in (9.33), for example, is written as

$$
p_{it1,w}^* = \frac{\exp(x_{it}'\beta + \theta + \sigma_\gamma\gamma_{iw}^*)}{[1+\exp(x_{it}'\beta + \theta + \sigma_\gamma\gamma_{iw}^*)]}.
$$

9.2.1.1.2 Formula for λ_{iut}

Conditional on γ_i^*, for $u < t$, the second-order expectation may be written as

$$E(Y_{iu}Y_{it}|\gamma_i^*) = \lambda_{iut}^*(\gamma_i^*) = \text{cov}(Y_{iu}, Y_{it}|\gamma_i^*) + \pi_{iu}^*\pi_{it}^* = \sigma_{iut}^* + \pi_{iu}^*\pi_{it}^*, \tag{9.34}$$

where by (7.149), the covariance between y_{iu} and y_{it}, conditional on γ_i^*, has the formula

$$\sigma_{iut}^* = \text{cov}(Y_{iu}, Y_{it}|\gamma_i^*) = \pi_{iu}^*(\gamma_i^*)(1-\pi_{iu}^*(\gamma_i^*))\Pi_{j=u+1}^t(p_{ij1}^* - p_{ij0}^*). \tag{9.35}$$

It then follows that the unconditional second-order raw moments have the formula

$$\lambda_{iut} = E(Y_{iu}Y_{it}) = M^{-1}\sum_{w=1}^M [\pi_{iu}^*(\gamma_{iw}^*)(1-\pi_{iu}^*(\gamma_{iw}^*)) \times \Pi_{j=u+1}^t(p_{ij1,w}^* - p_{ij0,w}^*) + \pi_{iu}^*(\gamma_{iw}^*)\pi_{it}^*(\gamma_{iw}^*)] . \tag{9.36}$$

Note that the first-order responses are used to construct ψ_{1i} for the β parameter and both squared and pairwise products are used to construct ψ_{3i} for the σ_γ^2 parameter. By the same token, to construct the basic distance function ψ_{2i} for the dynamic dependence parameter θ we use the pairwise products only. Thus,

$$\psi_{2i} = \sum_{u<t}^T [y_{iu}y_{it} - \lambda_{iut}]. \tag{9.37}$$

9.2.1.2 GMM Estimating Equation for $\alpha = (\beta',\ \theta,\ \sigma_\gamma^2)'$

By combining (9.31) and (9.37), for $\psi_i(y_i, \alpha) = [\psi_{1i}', \psi_{2i}, \psi_{3i}]'$, we now write a quadratic function as

$$Q_c(\alpha) = K^{-1}\left[\sum_{i=1}^K \psi_i(y_i,\alpha)\right]' C \left[\sum_{i=1}^K \psi_i(y_i,\alpha)\right], \tag{9.38}$$

[Hansen (1982)] where $\psi_i(y_i,\alpha)$ is the $(p+2)$-dimensional vector of moment functions corresponding to β, θ, and σ_γ^2, and C is a suitable weight matrix and must be positive definite. The GMM estimate of α is obtained by minimizing the quadratic function (9.38). To be specific, the GMM estimating equations for β, θ, and σ_γ^2 are given by

$$\frac{\partial \psi'}{\partial \alpha} C\psi = 0, \tag{9.39}$$

where $\psi = (\psi_1', \psi_2, \psi_3)'$ with

$$\psi_1 = K^{-1}\sum_{i=1}^K \psi_{1i},\ \ \psi_2 = K^{-1}\sum_{i=1}^K \psi_{2i},\ \ \psi_3 = K^{-1}\sum_{i=1}^K \psi_{3i},$$

and C is a weight matrix optimally chosen as

$$C = \left[K^{-2} \sum_{i=1}^{K} E\{\psi_i(y_i,\alpha)\psi_i'(y_i,\alpha)\} \right]^{-1}.$$

This C matrix is constructed in the next section.

9.2.1.2.1 Computation of the C Matrix

We now show how to compute the weight matrix C for the construction of the estimating equations given in (9.39). Remark that when one analyzes a semiparametric model, it becomes a challenge to construct an optimal weight matrix C. As a solution to this problem, Hansen (1982) suggested using a 'working' weight matrix C which may be constructed under certain relaxed conditions or parametric assumption. Under the present setup, one does not, however, need to use any 'working' C matrix. This is because the present dynamic binary mixed model (9.27) is completely specified and hence one can compute the optimal weight matrix C given by $C = [\text{cov}(\psi)]^{-1}$.

For convenience, we write the $\text{cov}(\psi)$ matrix under the present setup as

$$\text{cov}(\psi) = \begin{bmatrix} \text{var}(\psi_1) & \text{cov}(\psi_1,\psi_2) & \text{cov}(\psi_1,\psi_3) \\ & \text{var}(\psi_2) & \text{cov}(\psi_2,\psi_3) \\ & & \text{var}(\psi_3) \end{bmatrix}, \tag{9.40}$$

and provide the formulas for the components of this covariance matrix as follows.

First, we write the formula for the variance of ψ_1 which requires the unconditional moments of second order for the binary responses. To be specific,

$$\text{var}(\psi_1) = K^{-2} \sum_{i=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{T} \sigma_{iut} x_{it} x_{it}', \tag{9.41}$$

where the formulas for σ_{iut}, the variances and covariances of the repeated binary responses, are given by

$$\begin{aligned} \sigma_{itt} &= \pi_{it}[1-\pi_{it}] \\ \sigma_{iut} &= \lambda_{iut} - \pi_{iu}\pi_{it}, \end{aligned} \tag{9.42}$$

with λ_{iut} as in (9.36) and π_{it} as in (9.33).

Next, we write the formulas for the covariances requiring the moments of the data up to order three. These covariances are:

$$\text{cov}(\psi_1,\psi_2) = K^{-2} \left[\sum_{i=1}^{K} \sum_{u=1}^{T} \sum_{\ell \le t}^{T} x_{iu} \delta_{iu\ell t} - \sum_{i=1}^{K} \sum_{u=1}^{T} x_{iu}\pi_{iu} \sum_{i=1}^{K} \sum_{u \le t}^{T} \lambda_{iut} \right], \tag{9.43}$$

and

$$\mathrm{cov}(\psi_1,\psi_3) = K^{-2}\left[\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{\ell\le t}^{T} x_{iu}(\lambda_{iu\ell}+\delta_{iu\ell t}) - \sum_{i=1}^{K}\sum_{u=1}^{T} x_{iu}\pi_{iu} \times \sum_{i=1}^{K}\sum_{u\le t}^{T}(\pi_{iu}+\lambda_{iut})\right], \tag{9.44}$$

where the formulas for the raw second-order moments λ_{iut} are given in (9.36) and the raw third-order moments $\delta_{iu\ell}$ have the formulas given by

$$\delta_{iu\ell t} = E(Y_{iu}Y_{i\ell}Y_{it}) = M^{-1}\sum_{w=1}^{M}\sum_{y_{iu},y_{i\ell},y_{it}\notin s}^{1} \Pi_{j=2}^{T}\left[\tilde{p}_{ijy_{j-1}}(\gamma_{iw}^*)\tilde{p}_{i1}(\gamma_{iw}^*)\right]_{y_{iu}=1,y_{i\ell}=1,y_{it}=1}, \tag{9.45}$$

where

$$\tilde{p}_{i1}(\gamma_{iw}^*) = \exp\{y_{i1}(x_{i1}'\beta+\sigma_\gamma\gamma_{iw}^*)\}/[1+\exp(x_{i1}'\beta+\sigma_\gamma\gamma_{iw}^*)], \text{ and}$$
$$\tilde{p}_{ity_{i,t-1}}(\gamma_{iw}^*) = \exp\{y_{it}(x_{it}'\beta+\theta y_{i,t-1}+\sigma_\gamma\gamma_{iw}^*)\}/[1+\exp(x_{i1}'\beta+\theta y_{i,t-1}+\sigma_\gamma\gamma_{iw}^*)].$$

In (9.45), the sample space s contains the other $t-3$ elements out of all t elements $y_{i1},\ldots,y_{iu},\ldots,y_{i\ell},\ldots,y_{im},\ldots,y_{it}$.

The formulas for the remaining components contain the moments of the repeated responses up to order four. To be specific,

$$\mathrm{var}(\psi_2) = K^{-2}\left[\sum_{i=1}^{K}\sum_{u\le\ell}^{T}\sum_{m\le t}^{T}\phi_{iu\ell mt} - \left(\sum_{i=1}^{K}\sum_{u\le t}^{T}\lambda_{iut}\right)^2\right], \tag{9.46}$$

$$\mathrm{cov}(\psi_2,\psi_3) = K^{-2}\left[\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{\ell\le t}^{T}\delta_{iu\ell t} - \sum_{i=1}^{K}\sum_{u=1}^{T}\pi_{iu}\sum_{i=1}^{K}\sum_{u\le t}^{T}\lambda_{iut}\right] + \mathrm{var}(\psi_2), \tag{9.47}$$

and

$$\mathrm{var}(\psi_3) = K^{-2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{t=1}^{T}\sigma_{iut} + K^{-2}\left[\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{\ell\le t}^{T}\delta_{iu\ell t} - \sum_{i=1}^{K}\sum_{u=1}^{T}\mu_{iu}\sum_{i=1}^{K}\sum_{u\le t}^{T}\lambda_{iut}\right] + \mathrm{var}(\psi_2), \tag{9.48}$$

where $\phi_{iu\ell mt}$, the fourth-order unconditional uncorrected moments, have the formulas given by

$$\phi_{iu\ell mt} = E(Y_{iu}Y_{i\ell}Y_{im}Y_{it}) = M^{-1}\sum_{w=1}^{M}\sum_{y_{iu},y_{i\ell},y_{im},y_{it}\notin s}^{1} \Pi_{j=2}^{T}\left[\tilde{p}_{ijy_{i,j-1}}(\gamma_{iw}^*)\tilde{p}_{i1}(\gamma_{iw}^*)\right]_{y_{iu}=1,y_{i\ell=1},y_{im}=1,y_{it=1}}. \tag{9.49}$$

In (9.49), the sample space s contains the other $t-4$ elements out of all t elements $y_{i1},\ldots,y_{iu},\ldots,y_{i\ell},\ldots,y_{im},\ldots,y_{it}$. Note that as T is usually small in the panel data model, such as $T=4$ or more, and because y_{it}s $(t=1,\ldots,T)$ are binary, the third– and the fourth-order moments given by (9.45) and (9.49), respectively, are easily computed.

9.2.1.2.2 Computation of $\frac{\partial \psi'}{\partial \alpha}$

Now to solve (9.39) for β, θ, and σ_γ^2, we also need to compute the partial derivatives $\partial \psi'/\partial \alpha$. Note that ψ_1 and ψ_2 are functions of π_{it} and λ_{iut}, respectively, and ψ_3 is a function of both π_{it} and λ_{iut}. Thus, the computation of the derivative of the ψ function with regard to $\alpha=(\beta',\theta,\sigma_\gamma^2)'$ requires the computation of the derivatives $\partial \pi_{it}/\partial \beta_j$, $\partial \pi_{it}/\partial \theta$, $\partial \pi_{it}/\partial \sigma_\gamma^2$, and $\partial \lambda_{iut}/\partial \beta_j$, $\partial \lambda_{iut}/\partial \theta$ and $\partial \lambda_{iut}/\partial \sigma_\gamma^2$. For convenience, these derivatives are given in Exercises 9.1 and 9.2.

9.2.2 *GQL Estimation*

To construct the GQL estimating equations, follow Sutradhar (2003; 2004) and write a basic vector statistic containing the repeated responses and their distinct products for an individual. Let

$$u_i = (y_i', s_i')'$$

represent this vector with $y_i' = (y_{i1},\ldots,y_{iT})$ as the T-dimensional vector of responses for the ith individual and $s_i' = (y_{i1}y_{i2},\ldots,y_{iu}y_{it},\ldots,y_{i,T-1}y_{iT})'$ be the $(T-1)T/2$-dimensional vector of distinct pairwise products of the T responses. Let

$$\lambda_i = E(U_i) = [E(Y_i'),E(S_i')]'$$

be the expectation of the vector u_i, which is already computed in Section 9.2.1.1. To be specific, $E(Y_{it})=\mu_{it}$ and $E(Y_{iu}Y_{it})=\lambda_{iut}$ are known by (9.33) and (9.36), respectively. Furthermore, let Ω_i be the $\{T(T+1)/2 \times T(T+1)/2\}$ covariance matrix of u_i for the ith individual. In the GQL approach, one essentially minimizes the so-called generalized squared distance

$$\sum_{i=1}^{K}(u_i-\lambda_i)'\Omega_i^{-1}(u_i-\lambda_i) \qquad (9.50)$$

to estimate the parameters of the model, whereas the quadratic function $Q_c(\alpha)$ in (9.38) was minimized to obtain the GMM estimates. Once again it should be clear from (9.38) and (9.50) that in the GMM approach the quadratic distance function is written by using the distance between a combined statistic and its center (9.38), whereas in the GQL approach standardized distances for all individuals are combined to compute the generalized distance function (9.50). Note that minimization of the generalized squared distance (9.50) for the estimation of the $\alpha=(\beta',\theta,\sigma_\gamma^2)'$

parameter leads to the GQL estimating equations for $\alpha = (\beta', \theta, \sigma_\gamma^2)'$ as

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \alpha} \Omega_i^{-1} (u_i - \lambda_i) = 0, \tag{9.51}$$

which may be solved iteratively by using

$$\hat{\alpha}(r+1) = \hat{\alpha}(r) + \left(\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \alpha} \Omega_i^{-1} \frac{\partial \lambda_i}{\partial \alpha'} \right)_r^{-1} \left(\sum_{i=1}^{K} \frac{\partial \lambda_i'}{\partial \alpha} \Omega_i^{-1} (u_i - \lambda_i) \right)_r, \tag{9.52}$$

where $(\;)_r$ denotes that the quantity in the parenthesis is evaluated at $\alpha = \hat{\alpha}(r)$, the value of α obtained from rth iteration. Let $\hat{\alpha}_{GQL}$ denote the solution of (9.51) obtained by (9.52). This GQL estimator is consistent and it is more efficient than the GMM estimator. This is because the estimating equation (9.51) is constructed by using the true variance−covariance matrix of a basic statistic, whereas the GMM estimating equation (9.39) ignores the correlation structure of the data to form the combined basic statistics. An empirical study in Section 9.2.4 also confirms this superior relative efficiency performance of the GQL approach as compared to the GMM approach.

Note that to apply the iterative equation (9.52), one needs to compute the derivative vector $\partial \lambda_i' / \partial \alpha$, where

$$\lambda_i = [\pi_{i1}, \ldots, \pi_{it}, \ldots, \pi_{iT}, \lambda_{i12}, \ldots, \lambda_{iut}, \ldots, \lambda_{i(T-1)T}]'.$$

Also it is required to compute the $\Omega_i = \mathrm{cov}(U_i)$ matrix, where

$$u_i = (y_{i1}, \ldots, y_{iu}, \ldots, y_{iT}, y_{i1}y_{i2}, \ldots, y_{i\ell}y_{it}, \ldots, y_{i(T-1)}y_{iT})'.$$

But, as the derivatives of π_{it} and λ_{iut} with respect to $\alpha = (\beta', \theta, \sigma_\gamma^2)'$ are available from exercises 9.1 and 9.2, the vector of the derivatives, that is, $\partial \lambda_i' / \partial \alpha$, is known.

9.2.2.1 Computation of Ω_i

Now to compute the Ω_i matrix, we need to compute the variances $\mathrm{var}(Y_{iu})$ and $\mathrm{var}(Y_{i\ell}Y_{it})$ for all $u = 1, \ldots, T$, and $\ell < t, t = 2, \ldots, T$. Also, we need to compute the covariances $\mathrm{cov}(Y_{iu}, Y_{it})$, $\mathrm{cov}(Y_{iu}, Y_{i\ell}Y_{it})$ for $\ell < t$, and $\mathrm{cov}(Y_{iu}Y_{i\ell}, Y_{im}y_{it})$ for all possible values of $u < \ell$ and $m < t$. The formulas for some of these variances and covariances are already provided in Section 9.2.1.2.1 as basic properties of the model. To be specific, the formulas for the $\mathrm{var}(Y_{iu}) = \sigma_{iuu}$ and $\mathrm{cov}(Y_{iu}, Y_{it}) = \sigma_{iut}$ are given in (9.42). The remaining variances and covariances may be computed as follows. For example, the variance of the product variable $y_{i\ell}y_{it}$ is given by

$$\begin{aligned} \mathrm{var}(Y_{i\ell}Y_{it}) &= E(Y_{i\ell}^2 Y_{it}^2) - [E(Y_{i\ell}Y_{it})]^2 \\ &= E(Y_{i\ell}Y_{it})[1 - E(Y_{i\ell}Y_{it})] \end{aligned}$$

$$= \lambda_{i\ell t}[1-\lambda_{i\ell t}], \tag{9.53}$$

where the formula for $\lambda_{i\ell t}$ is known by (9.36). Similarly one obtains the other higher-order covariances. For example,

$$\text{cov}(Y_{iu}, Y_{i\ell}Y_{it}) = \delta_{iu\ell t} - \pi_{iu}\lambda_{i\ell t}, \tag{9.54}$$

where $\delta_{iu\ell t}$ is a third-order moment of the binary variables given by (9.45). Similarly, the $\text{cov}(Y_{iu}Y_{i\ell}, Y_{im}Y_{it})$ may be computed as

$$\begin{aligned}\text{cov}(Y_{iu}Y_{i\ell}, Y_{im}Y_{it}) &= E(Y_{iu}Y_{i\ell}Y_{im}Y_{it}) - E(Y_{iu}Y_{i\ell})E(Y_{im}Y_{it}) \\ &= \phi_{iu\ell mt} - \lambda_{iu\ell}\lambda_{imt},\end{aligned} \tag{9.55}$$

where the formula for the fourth moment $\phi_{iu\ell mt}$ is given in (9.49). This completes the construction of the $\{T(T+1)/2\} \times \{T(T+1)/2\}$ covariance matrix Ω_i.

9.2.3 *Efficiency Comparison: GMM Versus GQL*

9.2.3.1 Asymptotic Distribution of the GMM Estimator

Let $\hat{\alpha}_{GMM}$ be the GMM estimate of $\alpha = (\beta', \theta, \sigma_\gamma^2)'$ which is obtained by solving the GMM estimating equation (9.39). To be specific, this estimate is obtained by using the iterative equation

$$\hat{\alpha}_{GMM}(r+1) = \hat{\alpha}_{GMM}(r) + \left[\frac{\partial \psi'}{\partial \alpha} C \frac{\partial \psi}{\partial \alpha'}\right]_r^{-1} \left[\frac{\partial \psi'}{\partial \alpha} C \psi\right]_r, \tag{9.56}$$

where $[\ ]_r$ denotes that the quantity in the square bracket is evaluated at $\alpha = \hat{\alpha}_{GMM}(r)$, the value of α at the rth iteration. It then follows from (9.56) that asymptotically $(\text{as } K \to \infty)$

$$K^{\frac{1}{2}}(\hat{\alpha}_{GMM} - \alpha) \sim N\left[0, K\left(\frac{\partial \psi'}{\partial \alpha} C \frac{\partial \psi}{\partial \alpha'}\right)^{-1}\right]. \tag{9.57}$$

The normal distribution of the estimator follows from the fact that each of the components of the $\psi = (\psi_1', \psi_2, \psi_3)'$ vector is a sum of K independent quantities. More specifically, $\psi_1,\ \psi_2,\ \psi_3$ are constructed in (9.39) by using the sum of K independent quantities, namely, $f_{1i} = \sum_{t=1}^T x_{it}y_{it}$, $f_{2i} = \sum_{t=1}^T y_{it}$, and $f_{3i} = \sum_{u<t}^T y_{iu}y_{it}$. Consequently, normality follows from the multivariate central limit theorem [Mardia, Kent and Biby (1979, p. 51)]. See also Theorem 3.4 in Newey and McFadden (1993) for details on such asymptotic convergence.

9.2.3.2 Asymptotic Distribution of the GQL Estimator

Let $\hat{\alpha}_{GQL}$ be the GQL estimate of α which is obtained by solving the GQL estimating equation (9.51). It then follows that $\hat{\alpha}_{GQL}$ satisfies the equation $\sum_{i=1}^{I} M_i(\hat{\alpha}_{GQL}) = 0$, where $M_i(\alpha) = (\partial\lambda_i'/\partial\alpha)\Omega_i^{-1}(f_i - \lambda_i)$ so that $E[M_i(\alpha)] = 0$. It then follows, for example, from Theorem 3.4 of Newey and McFadden (1993) that with probability approaching 1, there is a unique solution, say $\hat{\alpha}_{GQL}$ to $\sum_{i=1}^{K} M_i(\alpha) = 0$ that satisfies

$$\sqrt{K}(\hat{\alpha}_{GQL} - \alpha) = -E\left[K^{-1}\sum_{i=1}^{K}\frac{\partial M_i(\alpha)}{\partial\alpha}\right]^{-1} K^{-1/2}\sum_{i=1}^{K} M_i(\alpha) + o_p(1), \quad (9.58)$$

implying the consistency of $\hat{\alpha}_{GQL}$ for α. Next, by the central limit theorem, it follows from (9.58) that as $K \to \infty$,

$$K^{\frac{1}{2}}(\hat{\alpha}_{GQL} - \alpha) \sim N\left[0, K\left[\sum_{i=1}^{K}\frac{\partial\lambda_i'}{\partial\alpha}\Omega_i^{-1}\frac{\partial\lambda_i}{\partial\alpha'}\right]^{-1}\right]. \quad (9.59)$$

In the next section, we report a comparative study from Sutradhar, Rao, and Pandit (2008) between the asymptotic variances of the GMM estimators computed by (9.57) and the asymptotic variances of the GQL estimators computed by (9.59). This asymptotic variance comparison was done through an empirical study based on a set of time-dependent covariates and a selected set of parameter values. In the following section, we discuss their simulation results on the small sample performances of the GMM estimates obtained from (9.56) and the GQL estimates obtained from (9.52).

9.2.3.3 Asymptotic Efficiency Comparison

It is clear from the last section that the GQL approach uses the true covariance structure of the model as the weight function in the estimating equation, whereas the GMM approach uses the moment equations that are constructed by ignoring the underlying correlation structure. This indicates that the GQL approach must produce estimates of the parameters with smaller standard errors as compared to the GMM estimators. In this subsection, we illustrate this efficiency gain of the GQL estimators over the GMM estimators by comparing their asymptotic variances numerically. In the next subsection we conduct a simulation study to examine the small sample performances of the GMM and GQL estimators.

For the asymptotic case, we compute the asymptotic variance−covariances of the GMM estimators of $\alpha = (\beta', \theta, \sigma_\gamma^2)'$ by (9.57) and those of the GQL estimators by (9.59). This we do under a binary panel data setup with $K = 500$ and $T = 4$. As far as the covariates are concerned, we choose two time-dependent covariates. The first covariate is considered to be:

$$x_{it1} = \begin{cases} 1 & \text{for } i = 1,\ldots,K/4;\ t = 1,2 \\ 0 & \text{for} i = 1,\ldots,K/4;\ t = 3,4 \\ -1 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 1 \\ 0 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 2,3 \\ 1 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 4 \\ t/T & \text{for } i = 3K/4+1,\ldots,I;\ t = 1,\ldots,4, \end{cases}.$$

whereas the second covariate is chosen to be

$$x_{it2} = \begin{cases} (t-2.5)/T & \text{for } i = 1,\ldots,K/2;\ t = 1,\ldots,T \\ 0 & \text{for } i = K/2+1,\ldots,K;\ t = 1,2 \\ 1 & \text{for } i = K/2+1,\ldots,K;\ t = 3,4. \end{cases}.$$

Furthermore, for true parameter values, we consider $\beta_1 = \beta_2 = 1.0$; $\theta = -0.3$, and $\sigma_\gamma^2 = 0.2$, 0.5, 0.8, 1.0, 1.2, 1.5, and 2.0. By using (9.57) and (9.59), we compute the asymptotic covariance matrices of

$$\hat{\alpha}_{GMM} = (\hat{\beta}_{1,GMM},\ \hat{\beta}_{2,GMM},\ \hat{\theta}_{GMM},\ \hat{\sigma}^2_{\gamma,GMM})',$$

and

$$\hat{\alpha}_{GQL} = (\hat{\beta}_{1,GQL},\ \hat{\beta}_{2,GQL},\ \hat{\theta}_{GQL},\ \hat{\sigma}^2_{\gamma,GQL})',$$

respectively. The diagonal elements of these covariance matrices, that is, the variances of these estimators, are presented in Table 9.1.

Table 9.1 Comparison of asymptotic variances (Var) of the GQL and GMM estimators for the estimation of the regression parameters (β_1 and β_2), the dynamic dependence parameter $\theta = -3.0$, and the variance component (σ_γ^2), of a logistic dynamic mixed model for binary panel data, with $T = 4$ and $K = 500$.

		Asymptotic Variances						
Method	Quantity	σ_γ^2= 0.2	0.5	0.8	1.0	1.2	1.5	2.0
GQL	Var($\hat{\beta}_1$)	0.010	0.010	0.010	0.010	0.011	0.011	0.013
	Var($\hat{\beta}_2$)	0.018	0.018	0.018	0.018	0.018	0.018	0.020
	Var($\hat{\theta}$)	0.035	0.035	0.036	0.036	0.037	0.038	0.041
	Var($\hat{\sigma}_\gamma^2$)	0.086	0.019	0.013	0.012	0.012	0.013	0.017
GMM	Var($\hat{\beta}_1$)	0.031	0.030	0.030	0.030	0.030	0.030	0.032
	Var($\hat{\beta}_2$)	0.051	0.046	0.041	0.038	0.036	0.034	0.034
	Var($\hat{\theta}$)	0.303	0.273	0.246	0.231	0.220	0.209	0.205
	Var($\hat{\sigma}_\gamma^2$)	0.349	0.075	0.048	0.044	0.043	0.047	0.061

It is clear from the table that the asymptotic variances under the GQL approach are uniformly smaller than those under the GMM approach. For example, when $\sigma_\gamma^2 = 1.0$, the GQL approach produces the asymptotic variances of the estimates of regression effects (β_1, β_2), dynamic dependence parameter (θ), and of the variance component of the random effects (σ_γ^2) as 0.010, 0.018, 0.036, 0.012, respectively, whereas the corresponding variances produced by the GMM approach are found to be 0.030, 0.038, 0.231, 0.044. To be more specific, the GQL estimates of β_1 and θ, for example, are, respectively, 3 and 28 times more efficient than the corresponding GMM estimates. This indicates that the GQL approach is definitely asymptotically more efficient as compared to the GMM approach.

Table 9.2 Comparison of simulated mean values, standard errors, and mean squared errors of the GQL and GMM estimates for the regression, dynamic dependence, and variance component parameters for $\theta = -1.0$ and selected values for σ_γ^2; $K = 100$; $T = 4$; true values of the regression parameters: $\beta_1 = \beta_2 = 1$; 500 simulations.

Variance			Estimates			
Component (σ_γ^2)	Method	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$	$\hat{\sigma}_\gamma^2$
0.50	GQL	Mean	1.033	1.195	−1.120	0.461
		SE	0.242	0.297	0.304	0.329
		MSE	0.060	0.127	0.107	0.102
	GMM	Mean	1.060	1.232	−1.173	0.532
		SE	0.314	0.402	0.570	0.392
		MSE	0.102	0.215	0.355	0.154
0.80	GQL	Mean	1.018	1.275	−1.155	0.743
		SE	0.238	0.318	0.319	0.292
		MSE	0.057	0.177	0.126	0.088
	GMM	Mean	1.046	1.295	−1.186	0.812
		SE	0.308	0.416	0.581	0.429
		MSE	0.097	0.260	0.372	0.184
1.00	GQL	Mean	1.008	1.305	−1.161	0.914
		SE	0.241	0.311	0.329	0.330
		MSE	0.058	0.189	0.134	0.116
	GMM	Mean	1.035	1.316	−1.178	0.983
		SE	0.324	0.407	0.586	0.468
		MSE	0.106	0.266	0.375	0.220
1.50	GQL	Mean	0.980	1.389	−1.154	1.310
		SE	0.242	0.333	0.343	0.242
		MSE	0.059	0.262	0.141	0.094
	GMM	Mean	1.015	1.438	−1.220	1.429
		SE	0.355	0.474	0.686	0.424
		MSE	0.127	0.416	0.519	0.185

9.2.3.4 Small Sample Efficiency Comparison: A Simulation Study

In order to examine the small sample performances of the GQL and GMM estimators, we carried out a simulation study with $K = 100$ clusters. Using $T = 4$

throughout, we considered the same data designs with two covariates as in the previous subsection. The true values of the regression parameters were considered to be $\beta_1 = \beta_2 = 1.0$. For the chosen design, we generated 500 simulated datasets under model (9.27) for a negative value of the dynamic dependence parameter, namely, $\theta = -1.0$, and four different values of σ_γ^2: 0.5, 0.8, 1.0, and 1.5. Then, for each group of 500 datasets associated with $\theta = -1.0$ and a chosen value for σ_γ^2 (there are four such groups of 500 datasets in total), we computed estimates for

$$\hat{\alpha} = (\hat{\beta}_1,\ \hat{\beta}_2,\ \hat{\theta},\ \text{and}\ \hat{\sigma}_\gamma^2)'$$

under the GMM approach by using (9.56), and under the GQL approach by using (9.52). For each of the two estimation approaches, these estimates were then used to compute their means (Mean) and standard errors (SE). We also computed the simulated mean squared errors (MSE) of the estimators of the four parameters under each of the two approaches. These simulated Mean, SE, and MSE are reported in Table 9.2 for the case when $\theta = -1.0$.

It appears from the table that the means of the GQL estimates for β_1,β_2, and θ appear to be closer to the true parameter values as compared to those of the GMM estimates. The GMM approach, however, appears to produce a slightly less biased estimate for the variance component parameter σ_γ^2. But the SE of the estimates for all four parameters are found to be smaller under the GQL approach as compared to the GMM approach. This in turn shows that the GQL approach always produces estimates with a smaller MSE than the GMM approach. For example, the results in Table 9.2 illustrate that when $\theta = -1.0$, and $\sigma_\gamma^2 = 1.00$, the estimates of the MSE for the GQL estimators are

$$0.058,\ 0.189,\ 0.134,\ 0.116,$$

versus

$$0.106,\ 0.266,\ 0.375,\ 0.220,$$

under the GMM approach.

In summary, both the asymptotic results of the previous subsection and the simulation results of this subsection clearly demonstrate the superiority of the GQL approach over the GMM approach in estimating the parameters of the dynamic binary mixed models.

9.2.4 *Fitting the Binary Dynamic Mixed Logit Model to the SLID data*

Recall that in Chapter 7, more specifically in Section 7.5.2, the Survey of Labour and Income Dynamics (SLID) data collected by Statistics Canada was analyzed by fitting the nonstationary AR(1) correlation model (7.70). As explained in Section 7.5.1, this study contains the longitudinal responses on employment status, that is, whether employed or unemployed for the full year, from $K = 15,731$ individuals for a period of four years from 1993 to 1996. Altogether the effects of 12 important covariates including gender, age group, region of residence, and education level, were computed by applying the GQL approach that accommodates the nonstationary longitudinal correlations. The results were reported in Table 7.6. However, if it is assumed that the mean response in a given year may be a function of the mean responses of the previous years, and also the response at a given year for an individual may be influenced by the individual's random effect, then it would be appropriate to fit the BDML model (9.27) to the data instead of fitting the linear dynamic conditional probability model (7.70). Becasue these assumptions are realistic for the SLID data, we now fit the BDML model to this dataset.

Note that with regard to fitting the BDML model (9.27) to the longitudinal data, in Sections 9.2.1 and 9.2.2, we have discussed the GMM and GQL estimation approaches, the GQL approach being more efficient as compared to the GMM approach. We now apply both procedures to the SLID data and estimate all 12 regression parameters, and dynamic dependence parameter (θ), as well as the variance of the individual random effects σ_γ^2.

Table 9.3 Estimates of regression and their estimated standard errors, as well as estimates and standard errors of dynamic dependence and variance component parameters.

	Estimation Method			
	GQL Approach		GMM Approach	
Parameters	Estimate	SE	Estimate	SE
Male vs Female (x_1)	−0.528	0.078	−0.536	0.094
Age group 2 vs 1 (x_2)	−1.525	0.035	−1.719	0.081
Age group 3 vs 1 (x_3)	−2.198	0.110	−2.108	0.169
Quebec vs Atlantic (x_4)	−0.728	0.072	−1.239	0.115
Ontario vs Atlantic (x_5)	−0.982	0.068	−1.326	0.101
Praries vs Atlantic (x_6)	−1.523	0.097	−2.001	0.131
BC & Alberta vs Atlantic (x_7)	−1.216	0.148	−1.913	0.187
Education medium vs low (x_8)	−1.572	0.039	−1.576	0.085
Education high vs low (x_9)	−2.326	0.149	−2.543	0.234
Marital status 2 vs 1 (x_{10})	0.189	0.082	0.251	0.136
Marital status 3 vs 1 (x_{11})	−0.616	0.223	−0.525	0.356
Marital status 4 vs 1 (x_{12})	−0.525	0.067	−0.629	0.144
θ	0.574	0.192	0.623	0.321
σ_γ	0.948	0.157	0.935	0.287

To be specific, we obtain the GMM estimates by using the GMM based iterative equation (9.56), and the GQL estimates by using the GQL based iterative equation (9.52). The standard errors of the GMM and GQL estimates are computed by using (9.57) and (9.59), respectively. The results are given in Table 9.3. It is clear from the table that the estimates for all parameters produced by both GMM and GQL approaches are close to each other. The SEs of the GQL estimates are, however, smaller than the GMM estimates. This is in agreement with the asymptotic and simulation results discussed in the last section. Consequently, it is sufficient to interpret the GQL estimates.

The GQL estimates for the dynamic dependence (θ) and variance component (σ_γ) parameters are found to be 0.574 and 0.948 with corresponding standard errors 0.192 and 0.157. The estimate for the dependence parameter indicates that the repeated binary responses are moderately positively correlated, $\theta = 0$ being the independence case. Also, the large value of $\sigma_\gamma = 0.948$ indicates that the unobservable random effects appear to have a large or moderately large influence on the mean and variance of the responses.

With regard to the BDML model based GQL regression effects, they are similar to those of the GQL estimates found in Table 7.6 by fitting the LDCP model to the data. It follows from Table 9.3 that the negative value -0.528 for the gender effect indicates that the male has a lower probability of an all-year unemployment as compared to the female. The negative values -1.525 and -2.918 of β_2 and β_3 indicate that the younger group has a higher probability of an all-year unemployment and the probability decreases for older age groups. As far as the effect of geographic location on all-year unemployment is concerned, it appears that the Prairies had the smallest probability of an all-year unemployment during 1993 to 1996 followed by BC, Ontario, Quebec, and Atlantic provinces. This follows from the fact that the regression estimates for Quebec, Ontario, BC, and Prairies are found to be -0.728, -0.982, -1.216, and -1.523, respectively. The larger negative value -2.326 for β_9 as compared to $\beta_8 = -1.572$ indicates that as the education level gets higher, the probability of an all-year unemployment gets smaller. Finally, with regard to the marital status, the positive value 0.189 for β_{10} means that the separated and divorced individuals have a higher probability of all-year unemployment as compared to the married and common-law spouse group. Similarly, the widowed had less probability of an all-year unemployment as compared to the single but never married individual.

9.2.5 GQL Versus Maximum Likelihood (ML) Estimation for BDML Model

The GQL estimation procedure has been discussed in Section 9.2.2, which was found to be better than the GMM approach in estimating the parameters of the underlying BDML model. Recall from Chapter 7, more specifically from Sections 7.7.2.2.2 and 7.7.2.2.3 that the ML approach and OGQL (optimal GQL) approaches were found to produce the same estimates for the parameters of the BDFL model. Note that the GQL in Section 9.2.2 is the same as the OGQL for the fixed model discussed in Section 7.7.2.2.2 as they both use the same first$-$ and second-order response based basic statistic. But the GQL in Section 9.2.2 and ML may not produce the same estimates for the BDML model. In this section, we report their comparative performances following Sutradhar, Bari, and Das (2010).

9.2.5.1 ML Estimation

For convenience we estimate $\alpha^* = (\beta',\ \theta,\ \sigma_\gamma)'$ by the ML approach, the GQL estimation for $\alpha = (\beta',\ \theta,\ \sigma_\gamma^2)'$ in Section 9.2.2 being easily adjustable for σ_γ. In the ML approach, one may obtain the estimate of $\alpha^* = (\beta',\ \theta,\ \sigma_\gamma)'$ by solving the likelihood estimating equation

$$\frac{\partial}{\partial \alpha^*} \log L = 0, \tag{9.60}$$

where the likelihood function in (9.29) may be written as

$$L(\beta,\ \theta,\ \sigma_\gamma) = \prod_{i=1}^{K} \int_{-\infty}^{\infty} p_{i10} \prod_{t=2}^{T} p_{ity_{i,t-1}} \phi(\gamma_i^*) d\gamma_i^*, \tag{9.61}$$

with $p_{i10} = (p_{i10}^*)^{y_{i1}}(1-p_{i10}^*)^{1-y_{i1}}$, and $p_{ity_{i,t-1}} = (p_{ity_{i,t-1}}^*)^{y_{it}}(1-p_{ity_{i,t-1}}^*)^{1-y_{it}}$, where $p_{ity_{i,t-1}}^*$ is defined in the model (9.27).

In (9.61), $\phi(\gamma_i^*)$ is the density function for standard normal γ_i^*. Note that one may solve (9.60) by using the iterative equation

$$\hat{\alpha}_{r+1}^* = \hat{\alpha}_r^* - \left[\left(\frac{\partial^2}{\partial \alpha^* \partial \alpha^{*\prime}} \log L\right)^{-1} \frac{\partial}{\partial \alpha^*} \log L\right]_{(r)}, \tag{9.62}$$

where $[\cdot]$ indicates that the quantity in the square bracket is evaluated at $\alpha^* = \hat{\alpha}_r^*$ obtained from the rth iteration. Further note that the computation of $\hat{\alpha}_{ML}$ by (9.62) and its variance will depend on the forms of p_{i10}^* and $p_{ity_{i,t-1}}^*$ which, under the present BDML model, by (9.27), are given as

$$p_{i10}^* = \frac{e^{x_{i1}'\beta + \sigma_\gamma \gamma_i^*}}{1 + e^{x_{i1}'\beta + \sigma_\gamma \gamma_i^*}} \quad \text{and} \quad p_{ity_{i,t-1}}^* = \frac{e^{x_{it}'\beta + \theta y_{it-1} + \sigma_\gamma \gamma_i^*}}{1 + e^{x_{it}'\beta + \theta y_{it-1} + \sigma_\gamma \gamma_i^*}}.$$

Unlike the GQL estimation discussed in Section 9.2.2, the computations for the ML estimation are complex and lengthy. For example, the components of $\partial \log L / \partial \alpha^*$ for (9.60) under the dynamic logit mixed model have the forms:

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{K} \sum_{t=1}^{T} \left[y_{it} - \frac{A_{it}}{J_i}\right] x_{it}, \tag{9.63}$$

$$\frac{\partial \log L}{\partial \theta} = \sum_{i=1}^{K} \sum_{t=1}^{T} \left[y_{it} - \frac{A_{it}}{J_i}\right] y_{it-1}, \tag{9.64}$$

and

$$\frac{\partial \log L}{\partial \sigma_\gamma} = \sum_{i=1}^{K} \frac{M_i}{J_i}, \tag{9.65}$$

respectively, where

$$J_i = \int_{-\infty}^{\infty} \exp(\sigma_\gamma \gamma_i^* s_i) \Delta_i \phi(\gamma_i^*) d\gamma_i^*,$$

with $s_i = \sum_{t=1}^{T} y_{it}$ and $\Delta_i = \left[\prod_{i=1}^{T} \{1 + \exp(x_{it}'\beta + \theta y_{it-1} + \sigma_\gamma \gamma_i^*)\}\right]^{-1}$, and A_{it} and M_i are given as

$$A_{it} = \int_{-\infty}^{\infty} \exp(\sigma_\gamma \gamma_i^* s_i) \Delta_i p^*_{ity_{i,t-1}} \phi(\gamma_i^*) d\gamma_i^*,$$

$$M_i = \int_{-\infty}^{\infty} \exp(\sigma_\gamma \gamma_i^* s_i) \left[\sum_{t=1}^{T} (y_{it} - p^*_{ity_{i,t-1}})\right] \Delta_i \gamma_i^* \phi(\gamma_i^*) d\gamma_i^*,$$

respectively.

Let $\hat{\alpha}^*_{ML}$ be the solution obtained from (9.62). One may then show that as $K \to \infty$, $\hat{\alpha}^*_{ML}$ has asymptotic normal distribution given by

$$\sqrt{K}(\hat{\alpha}^*_{ML} - \alpha^*) \sim \mathrm{N}(0, KV_{ML}^{-1}) \tag{9.66}$$

[Amemiya (eqns. (11.1.38), 1985); Gourieroux and Monfort (1981)] where $V_{ML} = -E\left\{(\partial^2/\partial\alpha^*\partial\alpha^{*\prime})\log L\right\}$. The exact computation for this covariance matrix is not possible under the present dynamic mixed model, but, it can be computed numerically, by simulating the random effects γ_i^* (Sutradhar, 2004).

9.2.5.2 Relative Performances of the GQL and ML Approaches for BDML model: A Simulation Study

To examine the relative performance of the ML approach as compared to the GQL approach in estimating all parameters β, θ, and σ_γ, for the BDML model (9.27), we now carry out a Monte Carlo study based on 1000 simulations. As far as the simulation design is concerned, we consider $T = 4$ repeated binary responses from each of $K = 100$ independent individuals. To represent the dynamic dependence in repeated responses, the data are generated with lag 1 dependence parameter $\theta = 1.0$. With regard to the dimension of the regression effects, we consider $p = 2$, with $\beta_1 = \beta_2 = 0.0$. As far as the covariates are concerned, we choose the first covariate as

$$x_{it1} = \begin{cases} 1/2 & \text{for } i = 1,\ldots,K/4;\ t = 1,2 \\ 0 & \text{for } i = 1,\ldots,K/4;\ t = 3,4 \\ -1/2 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 1 \\ 0 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 2,3 \\ 1/2 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 4 \\ t/(2T) & \text{for } i = 3K/4+1,\ldots,K;\ t = 1,\ldots,4, \end{cases}$$

whereas the second covariate is chosen as

$$Pr(x_{it2} = 1) = \begin{cases} 0.3 & \text{for } t = 1 \\ 0.5 & \text{for } t = 2,3 \\ 0.8 & \text{for } t = 4. \end{cases}$$

Note that even though x_{it2}s are generated from the binary distributions, they were kept the same under all simulations. Thus, this x_{it2} is also a fixed covariate. As far as the selection of the additional parameter σ_γ^2 is concerned, we choose both small and large values, namely, $\sigma_\gamma = 0.5$, 0.8, and 1.2.

Next, in each of the 1000 simulations, we solve the estimating equation in (9.52) under the BDML model to obtain the GQL estimates, and use (9.62) to obtain the ML estimates. We then obtain the simulated means (SMs) and simulated variances (SVs) of these 1000 estimates for each parameter. Note that as opposed to the fixed models considered in Chapter 7, it is expected that the ML estimates would be different in general as compared to the GQL estimates under the mixed models. Also it is anticipated that the estimates may be much more biased under the mixed models as compared to the fixed models. Because an estimate with large bias and small standard error becomes inconsistent, along with the SMs and SVs, in this section, we compute the simulated relative biases (SRBs) of the estimates as opposed to their simulated mean squared errors (SMSEs). The percentage relative biases, for example, the percentage simulated relative bias (SRB) of $\hat{\theta}_{ML}$ is defined by

$$SRB(\hat{\theta}_{ML}) = \frac{|\mathrm{SM}(\hat{\theta}_{ML}) - \theta_0|}{\sqrt{\mathrm{SV}(\hat{\theta}_{ML})}} \times 100.$$

The SMs, SVs, and SRBs of the GQL and ML estimates for $\theta = 1$, $\beta_1 = \beta_2 = 0$, and for selected values of the other parameter σ_γ, under the BDML model, are shown in Table 9.4.

Table 9.4 Simulated mean, variance, and relative bias of the ML and GQL estimates under the BDML model for $\beta_1 = \beta_2 = 0.0$ and $\theta = 1.0$.

		Method of Estimation							
		ML				GQL			
σ_γ		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$	$\hat{\sigma}_\gamma$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$	$\hat{\sigma}_\gamma$
0.5	SM	−0.021	0.058	1.000	0.390	−0.021	0.078	0.973	0.503
	SV	0.137	0.041	0.055	0.075	0.140	0.041	0.071	0.308
	SRB	6	29	0	40	6	39	10	1
0.8	SM	−0.049	0.066	1.029	0.653	−0.037	0.095	0.984	0.774
	SV	0.149	0.045	0.061	0.132	0.151	0.045	0.065	0.214
	SRB	13	31	12	40	10	45	6	6
1.2	SM	−0.113	0.034	1.121	0.878	−0.0404	0.117	0.990	1.209
	SV	0.178	0.066	0.099	0.364	0.179	0.053	0.077	0.083
	SRB	27	13	39	53	10	51	4	3

The results in Table 9.4 show that irrespective of the true value of σ_γ, small or large, the ML approach always produces a σ_γ estimate with larger RB than the GQL approach. With regard to the estimation of $\theta = 1$, the ML approach produces estimates with less RBs when σ_γ is small, but it produces estimates with larger RBs than the GQL approach when σ_γ is large such as $\sigma_\gamma = 0.8,\ 1.2$. For the estimation of the regression effects, the ML and GQL approaches exhibit mixed performances. For example, when $\sigma_\gamma = 0.5$, they perform almost the same in estimating β_1 and β_2; and for large σ_γ, the ML approach estimates β_2 with smaller RBs but estimates β_1 with larger RBs, as compared to the GQL approach. Thus, the results of Table 9.4 indicate that in general the GQL approach performs better than the ML approach in estimating the parameters of the BDML model. Note, however, that because the results of Table 9.4 show that a selected true parameter value falls in the interval: estimate minus/plus two times standard deviation, both ML and GQL approaches clearly produce consistent estimates, which is in agreement with the asymptotic results given in (9.66) for the ML estimators and in (9.58) for the GQL estimators.

Note that in general when ML estimates are found to be unbiased or almost unbiased, they are recommended in practice as compared to other competitors such as the method of moments (MM) or QL estimates. This is because, the standard errors of the ML estimates are usually found to be less than those of the other estimates, leading the ML estimates to be consistent and more efficient as well. In such cases, the mean squared errors of the ML estimates will also be smaller as compared to the other estimates. It is, however, well known that ML estimates can be more biased for certain parameters such as for the variance parameter of the well-known normal distribution as compared to the MM or QL estimates. In the present BDML setup, the variance parameter of the random effects plays a complicated role to interpret the variation and other moments of the binary data. Thus, it was not surprising to observe that the ML estimates become biased for such variance and dynamic dependence parameters. By the same token, when ML estimates are biased, the traditional MSE comparison may not reveal the real properties of the estimates. This is because a biased estimate with smaller standard error will mostly converge to a wrong place, that is, to a value different than the true value of the parameter. This is why, in the aforementioned discussion, we have concentrated on the RB instead of MSE to compare the actual performance of the estimates in the sense of their convergence to the true values.

9.3 A Binary Dynamic Mixed Probit (BDMP) Model

In binary panel data analysis, there are situations where the binary outcomes are thought to arise from a standard normal latent process [Amemiya (1985)]. To be specific, let the binary observation y_{it} be obtained from the standard normal distribution of y^*_{it} as follows.

$$y_{i1} = \begin{cases} 1 \text{ if } y_{i1}^* < x_{i1}'\beta + \sigma_\gamma \gamma_i^* \\ 0 \text{ otherwise,} \end{cases} \tag{9.67}$$

and for $t = 2, \ldots, T$,

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* < x_{it}'\beta + \theta y_{i,t-1} + \sigma_\gamma \gamma_i \\ 0 & \text{otherwise.} \end{cases} \tag{9.68}$$

One may then write the binary probabilities as

$$Pr(y_{it} = 1|\gamma_i) = \begin{cases} \Phi(x_{i1}'\beta + \sigma_\gamma \gamma_i^*) = p_{i10}^*, & \text{for } i = 1, \ldots, K;\ t = 1 \\ \Phi(x_{it}'\beta + \theta y_{i,t-1} + \sigma_\gamma \gamma_i^*), & \text{for } i = 1, \ldots, K;\ t = 2, \ldots, T \\ = p_{ity_{i,t-1}}^* \end{cases}$$

$$= F_{it}, \text{ say,} \tag{9.69}$$

where $\Phi(\cdot)$ is the cumulative probability of the standard normal variable. Note that as opposed to the BDML model (9.27), the binary model in (9.69) is known as the binary dynamic mixed probit model that arises from standard normal latent process, whereas as indicated in exercise 9.3, the BDML model arises from logistic latent distribution.

Further note that because the ML approach performed worse than the GQL approach under the BDML model, and because of the fact that the GQL approach is simpler than the ML approach under any model, we do not attempt any study for the comparison of the ML and GQL approaches for the estimation of the parameters of the probit model. Thus, in the next section we examine the performance of the GQL approach only, for the estimation of the parameters of the BDMP model. We also show the performance of the GQL approach for the same parameter values under the BDML model.

9.3.1 GQL Estimation for BDMP Model

The GQL estimating equation for $\alpha = [\beta'.\ \theta,\ \sigma_\gamma^2]'$ under the BDML model (9.27) is given by (9.51). The GQL estimating equation for $\alpha^* = [\beta'.\ \theta,\ \sigma_\gamma]'$ under the BDMP model (9.69) has similar form as that of (9.51), but

$$\lambda_i = E[U_i],\ \ \Omega_i = \text{cov}[U_i],\ \text{and}\ \frac{\partial \lambda_i'}{\partial \alpha^*},$$

have to be computed now under the probit model (9.69). Let

$$\lambda_{i,P}\ \text{and}\ \Omega_{i,P}$$

denote the expectation and the covariance matrix of u_i under the probit (P) model (9.69). These moments may be computed following (9.33) and (9.36), respectively, but by using the formulas for p^*_{i10} and $p^*_{ity_{i,t-1}}$ from (9.69) instead of (9.27). We now write the GQL estimating equation for α^* as

$$\sum_{i=1}^{K} \frac{\partial \lambda'_{iP}}{\partial \alpha^*} \Omega_{i,P}^{-1} (u_i - \lambda_{i,P}) = 0, \qquad (9.70)$$

where the first-order derivatives have the formulas as in Exercise 9.3. The estimators are consistent and highly efficient. Also, the GQL estimator $\hat{\alpha}^*_{GQL,P}$ obtained from (9.70) has the asymptotic $(K \to \infty)$ normal distribution with mean α^* and covariance matrix

$$\mathrm{cov}[\hat{\alpha}^*_{GQL,P}] = \left[\sum_{i=1}^{K} \frac{\partial \lambda'_{iP}}{\partial \alpha^*} \Omega_{i,P}^{-1} \frac{\partial \lambda_{iP}}{\partial \alpha^{*\prime}} \right]^{-1}. \qquad (9.71)$$

In the next section, we report some simulation results from Sutradhar, Bari, and Das (2010) on the finite sample performance of the GQL estimation approach for the parameters of the BDMP model.

9.3.2 GQL Estimation Performance for BDMP Model: A Simulation Study

For the simulation study, we consider the same two covariates as in Section 9.2.5.2. Also, even though we have examined the performance of the GQL approach for the estimation of the parameters of the BDML model through a simulation study in Section 9.2.5.2 (see Table 9.4), we include the estimation for this model in the current simulation study. For the regression parameters we choose $\beta_1 = \beta_2 = 0$, the same as in the other simulation study in Section 9.2.5.2, but consider new values for θ, namely $\theta = 0.0, 2.0$. Three different values for σ_γ are considered, namely $\sigma_\gamma = 0.5, 0.8, 1.2$. The GQL estimates for parameters of the BDMP and BDML models are obtained by solving (9.70) and (9.51), respectively. The asymptotic estimated variances (AEV) of the GQL estimators under BDMP and BDML models are computed by using (9.71) and (9.59), respectively. Based on 1000 simulations, we display the SM, SV, SRB, and AEV of the GQL estimates of all three parameters in Table 9.5.

It is clear from Table 9.5 that under both probit and logit mixed models, the SMs in general appear to agree with the coresponding true values of the parameters. The GQL estimates of β_1 appear to be much better in the sense of RB, as compared to those of β_2, under both probit and logit models. The GQL approach appears to perform quite well in estimating the variance of the random effects under the logit mixed model. The GQL estimates for this variance parameter appear to have larger relative biases under the probit mixed model as compared to the logit model. With regard to the estimation of the dynamic dependence parameter θ, the GQL approach

Table 9.5 Simulated mean, variance, and relative bias, and asymptotic estimate of variance, based on the GQL approach under the BDPM and BDLM models; 1000 simulations, $\beta_1 = \beta_2 = 0.0$.

			Model							
			Probit				Logit			
θ	σ_γ	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$	$\hat{\sigma}_\gamma$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$	$\hat{\sigma}_\gamma$
0.0	0.5	SM	0.018	0.008	0.017	0.447	−0.014	0.080	−0.041	0.505
		SV	0.050	0.013	0.026	0.028	0.122	0.035	0.062	0.140
		AEV	0.048	0.015	0.028	0.023	0.115	0.035	0.064	0.123
		SRB	8	7	11	32	4	43	17	1
	0.8	SM	0.013	0.050	0.006	0.748	−0.032	0.097	−0.021	0.790
		SV	0.060	0.016	0.030	0.021	0.135	0.036	0.067	0.067
		AEV	0.057	0.018	0.034	0.028	0.129	0.040	0.074	0.065
		SRB	5	40	3	39	9	51	8	4
	1.2	SM	0.011	0.055	0.018	1.107	−0.043	0.116	−0.014	1.213
		SV	0.071	0.020	0.040	0.030	0.162	0.042	0.081	0.070
		AEV	0.070	0.024	0.044	0.042	0.151	0.048	0.090	0.076
		SRB	4	39	9	54	11	57	5	5
2.0	0.5	SM	−0.007	0.055	2.063	0.470	−0.002	0.077	2.010	0.459
		SV	0.078	0.029	0.116	0.048	0.164	0.057	0.090	0.542
		AEV	0.074	0.030	0.133	0.044	0.149	0.055	0.081	0.509
		SRB	3	32	19	14	1	32	3	6
	0.8	SM	−0.010	−0.016	2.095	0.698	−0.030	0.092	2.016	0.773
		SV	0.085	0.032	0.092	0.049	0.168	0.063	0.094	0.124
		AEV	0.089	0.035	0.103	0.051	0.167	0.061	0.092	0.102
		SRB	3	9	31	46	7	37	5	8
	1.2	SM	−0.018	0.061	2.074	1.059	−0.061	0.110	2.011	1.195
		SV	0.098	0.033	0.136	0.037	0.205	0.072	0.108	0.111
		AEV	0.115	0.045	0.149	0.046	0.197	0.072	0.112	0.114
		SRB	6	34	20	73	14	41	3	2

performs well under both models when $\theta = 0$, that is, for the longitudinally independent case, but for large $\theta = 2.0$, the GQL approach appears to produce better estimates under the logit model than the probit model. In summary, the GQL approach appears to perform quite well in estimating all parameters of the dynamic probit and logistic mixed models. Furthermore, this approach is simpler as compared to the ML approach, especially under the probit mixed model.

9.3.2.1 Random Effects Mis-specification: True t Versus Working Normal Distributions For Random Effects

In practice, it is standard to assume that the random effects involved in the mixed models follow a Gaussian distribution. The simulation results with regard to the performance of the GQL estimation for normal random effects based probit and logit mixed models were shown in Table 9.5. In this section, we conduct an additional small simulation study by generating random effects γ_i^* for $i = 1, \ldots, K$ with $K = 100$ from a heavy-tailed t-distribution with degrees of freedom $\nu = 6.55$, and examine the robustness of the normal random effects based GQL approach. Note that $\nu = 6.55$ yields the variance component $\sigma_\gamma = \sqrt{\nu/(\nu-2)} = 1.2$. As far as the covariates and parameter values are concerned, we consider the same covariates and parameter values as in the previous Section 9.3.2. The SMs, SVs, and SRBs of the normality assumption based GQL estimates for all parameters β_1, β_2, θ, and σ_γ under both probit and logit mixed models are reported in Table 9.6.

Table 9.6 Simulated mean, standard error, and relative bias of the GQL estimates computed by using the normal random effects based BDPM and BDLM models when data are generated from the BDPM and BDLM models with the random effects following t-distribution with degrees of freedom $\nu = 6.55$ ($\sigma_\gamma = 1.2$); $\beta_1 = \beta_2 = 0.0$; based on 1000 simulations.

		Model							
		Probit				Logit			
θ	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$	$\hat{\sigma}_\gamma$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$	$\hat{\sigma}_\gamma$
0.0	SM	−0.101	0.016	−0.066	1.253	−0.139	0.036	−0.105	1.280
	SV	0.069	0.021	0.044	0.030	0.161	0.042	0.091	0.066
	SRB	39	11	32	31	35	18	35	31
2.0	SM	−0.166	−0.050	2.099	1.199	−0.143	0.012	1.969	1.300
	SV	0.086	0.034	0.124	0.036	0.203	0.070	0.116	0.102
	SRB	57	27	28	1	32	5	9	31

When these results of Table 9.6 are compared with those in Table 9.5 for $\sigma_\gamma = 1.2$ (bottom four rows from each half), the estimate of β_1 appears to be affected adversely under both probit and logit mixed models. The estimate of σ_γ also appears to be affected but under the logit model only. For example, Table 9.6 shows that the SRB for σ_γ estimate is 31, when $\theta = 0$, 2.0 under the logit model, whereas true normality based GQL approach (Table 9.5) produces SRBs 5 and 2 for θ values 0 and 2.0, respectively. Thus, in general, the GQL approach appears to be sensitive to the correct distributional assumption for the random effects.

Note that in the linear model set up, the moments based estimation approaches such as the present GQL approach are known to be less affected by the departure of the distributional assumption from normality. This is especially true when the error distribution for a linear model remain symmetric but different from the normal distribution. In the present nonlinear setup, it is, however, expected that this stability property may not hold. This is mainly because when the distribution of the random effects in the exponent of the probability function is replaced by another nonnormal symmetric or asymmetric distribution, the distributional properties, that is, the moments of the actual data get changed in a complicated way. Thus, the results in Table 9.6 showing that a change in the distribution of the random effects may affect the estimates considerably, appear to be reasonable. However, it is practical to consider that the random effects follow the Gaussian distribution.

Exercises

9.1. (Section 9.2.1.2.2) [First-order derivatives of π_{it} with respect to $\alpha^* = (\beta,\ \theta,\ \sigma_\gamma)'$ under the logit model (9.27)]
Write the binary probabilities in the BDML model (9.27) as

$$F_L(z^*) = \frac{e^{z^*}}{1+e^{z^*}}, \tag{9.72}$$

[Cumulative logistic distribution function] so that

$$f_L(z^*) = \frac{\partial F_L(z^*)}{\partial z^*} = F_L(z^*)(1-F_L(z^*)). \tag{9.73}$$

Now using $f_L(z^*)$, verify from (9.33) that the derivatives of π_{it} with respect to $\beta,\ \theta,$ and σ_γ have the formulas given by:

$$\frac{\partial \pi_{it}}{\partial \alpha^*} = \frac{1}{M}\sum_{w=1}^{M} \frac{\partial \pi^*_{it}(\gamma^*_{iw})}{\partial \alpha^*}, \tag{9.74}$$

where, for

$$\alpha^* = (\beta',\ \theta,\ \sigma)' \equiv ({\alpha^*}'_1,\ \alpha^*_2,\ \alpha^*_3)'$$

and by using the recurrence relationship (9.33), one obtains

$$\frac{\partial \pi^*_{it}(\gamma^*_{iw})}{\partial \alpha^*_v} = c_{v,t} + d_{v,t}\pi^*_{i,t-1} + [p^*_{it1} - p^*_{it0}]\,\frac{\partial \pi^*_{i,t-1}(\gamma^*_{iw})}{\partial \alpha^*_v}, \tag{9.75}$$

with

$$c_{v,t} = \begin{cases} f_L\left(x'_{it}\beta + \sigma_\gamma\gamma^*_{iw}\right)\ x_{it} & \text{for } v=1 \\ 0 & \text{for } v=2 \\ f_L\left(x'_{it}\beta + \sigma_\gamma\gamma^*_{iw}\right)\ \gamma^*_{iw} & \text{for } v=3, \end{cases} \tag{9.76}$$

and

$$d_{v,t} = \begin{cases} [f_L(x'_{it}\beta + \theta + \sigma_\gamma\gamma^*_{iw}) - f_L(x'_{it}\beta + \sigma_\gamma\gamma^*_{iw})]x_{it} & \text{for } v=1 \\ f_L\left(x'_{it}\beta + \theta + \sigma_\gamma\gamma^*_{iw}\right) & \text{for } v=2 \\ \left[f_L(x'_{it}\beta + \theta + \sigma_\gamma\gamma^*_{iw}) - f_L(x'_{it}\beta + \sigma_\gamma\gamma^*_{iw})\right]\gamma^*_{iw} & \text{for } v=3. \end{cases} \tag{9.77}$$

9.2. (Section 9.2.1.2.2) [First-order derivatives of λ_{iut} with respect to $\alpha^* = (\beta,\ \theta,\ \sigma_\gamma)'$ under the logit model (9.27)]
Use the notation from Exercise 9.1 and verify from (9.36) that the derivatives of λ_{iut} with respect to $\beta,\ \theta,$ and σ_γ have the formulas given by:

$$\frac{\partial \lambda_{iut}}{\partial \alpha^*} = \frac{1}{M}\sum_{w=1}^{M} \frac{\partial \lambda^*_{iut}(\gamma^*_{iw})}{\partial \alpha^*}, \tag{9.78}$$

where by (9.34),

$$\lambda^*_{iut} = \pi^*_{iu}(1-\pi^*_{iu}) \prod_{j=u+1}^{t} \left[p^*_{ij1} - p^*_{ij0}\right] + \pi^*_{iu}\pi^*_{it},$$

so that for

$$\alpha^* = (\beta', \theta, \sigma)' \equiv ({\alpha^*}'_1,\ \alpha^*_2, \alpha^*_3)',$$

one obtains

$$\begin{aligned}\frac{\partial \lambda^*_{iut}}{\partial \alpha^*_v} &= (1-2\pi^*_{iu})\frac{\partial \pi^*_{iu}}{\partial \alpha^*_v} \prod_{j=u+1}^{t} \left(p^*_{ij1} - p^*_{ij0}\right) \\ &+ [\pi^*_{iu}(1-\pi^*_{iu})] \sum_{j=u+1}^{t} d_{v,j} \prod_{l\neq j, l=u+1}^{t} (p^*_{il1} - p^*_{il0}) \\ &+ \left[\pi^*_{it}\frac{\partial \pi^*_{iu}}{\partial \alpha^*_v} + \pi^*_{iu}\frac{\partial \pi^*_{it}}{\partial \alpha^*_v}\right].\end{aligned} \tag{9.79}$$

9.3. (Section 9.3.1) [First-order derivatives of π_{it} and λ_{iut} with respect to $\alpha^* = (\beta',\ \theta,\ \sigma_\gamma)'$ under the probit model (9.69)]
Justify that the first-order derivatives

$$\frac{\partial \pi_{it}}{\partial \alpha^*} \text{ and } \frac{\partial \lambda_{iut}}{\partial \alpha^*},$$

under the BDMP model (9.69) may be computed by replacing the logit (L) model based $f_L(z^*)$ in Exercises 9.1 and 9.2, with the probit model based $f_N(z^*)$ which has the formula given by

$$f_N(z^*) = \frac{\partial}{\partial z^*}\left[\int_{-\infty}^{z^*} \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}z^2]\partial z\right] = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}{z^*}^2], \tag{9.80}$$

a standard normal density.

References

1. Amemiya, T. (1985). *Advanced Econometrics,* Cambridge, MA: Harvard University Press.

2. Gourieroux, C. & Monfort, A. (1981). Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. *J. Econometrics,* 17, 83 – 91.
3. Hansen, L.-P. (1982). Large sample properties of generalized method of moment estimators. *Econometrica,* 50, 1029 – 1054.
4. Heckman, J. J. (1981). Statistical models for discrete panel data, in *Structural Analysis of Discrete Data with Econometric Applications,* ed. by D. McFadden and C. F. Manski. Cambridge, MA: MIT Press.
5. Honore, B. E. & Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica,* 68, 839 – 874.
6. Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.,* 93, 720 – 729.
7. Jiang, J. & Zhang, W. (2001). Robust estimation in generalized linear mixed models. *Biometrika,* 88, 753 – 765.
8. Manski, C. F. (1987). Semi-parametric analysis of random effects linear models from binary panel data. *Econometrica,* **55,** 357 – 362.
9. Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press Inc.
10. Newey, W. K. & McFadden, D. (1993). Estimation in large samples, in *Handbook of Economics*, Vol. 4, eds. D. McFadden and R. Engler, Amsterdam: North Holland.
11. Sutradhar, B. C. (2003). An Overview on Regression Models for Discrete Longitudinal Responses. *Statist. Sci.,* 18, 377 – 393.
12. Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhya B,* 66, 261 – 289.
13. Sutradhar, B. C., Bari, W. & Das, K. (2010). On probit versus logit dynamic mixed models for binary panel data. *J. Statist. Comput. Simul.,* 80, 421 – 441.
14. Sutradhar, B. C. & Farrell, P. J. (2004). On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data. *Sankhya B,* 69, 448 – 467.
15. Sutradhar, B. C. & Jowaheer, V. (2003). On Familial Longitudinal Poisson Mixed Models with Gamma Random Effects. *J. Multivariate Anal.,* 87, 398 – 412.
16. Sutradhar, B. C., Rao, R. P,& Pandit, V. N. (2008). Generalized method of moments versus generalized quasi-likelihood inferences in binary panel data models. *Sankhya B,* 70, 34 – 62.

Chapter 10
Familial Longitudinal Models for Count Data

In Chapter 4, we discussed familial models for count data, where count responses along with a set of covariates are collected from the members of a large number of independent families. In Chapter 6, we discussed longitudinal models for count data, where count responses along with a set of covariates are collected from a large number of independent individuals over a small period of time. In practice there are situations where the count responses and their corresponding covariates are collected in a familial longitudinal setup. In this setup, count responses and the associated covariates are collected from the members of a large number of independent families over a small period of time. For example, in health care utilization data, the number of visits to the physician by the members of a large number of independent families may be recorded over a period of several years. Also the information on the covariates: gender, number of chronic conditions, education level, and age may be recorded for the members of each family. To analyze this type of familial longitudinal data, one needs to combine the familial and longitudinal models from Chapters 4 and 6, and construct a general familial longitudinal model. As expected, the count responses, in such a setup, will exhibit a familial longitudinal correlation structure. The purpose of this chapter is to take this two-way correlation structure into account, and develop a suitable estimation approach, such as generalized quasi-likelihood (GQL) estimation approach, for the estimation of the regression effects and the familial correlation index parameter, whereas the longitudinal correlation parameter is estimated by using the well-known method of moments.

10.1 An Autocorrelation Class of Familial Longitudinal Models

Let y_{ijt} denote the count response for the jth $(j = 1,\ldots,n_i)$ individual on the ith $(i = 1,\ldots,K)$ family/cluster at a given time t $(t = 1,\ldots,T)$. Also, let $x_{ijt} = (x_{ijt1},\ldots,x_{ijtp})'$ denote the p covariates associated with the response y_{ijt}, and β denotes the effect of the covariate vector x_{ijt} on y_{ijt}. Note that as the members of the ith $(i = 1,\ldots,K)$ family are likely to be influenced by a common family effect, say

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_10,

γ_i, the count responses of any two members of the same family at a given time are likely to be correlated. This correlation is referred to as the familial correlation. Furthermore, conditional on the unobservable family effect γ_i, the repeated count data collected from the same member of the ith family are also likely to be correlated. This correlation is referred to as the longitudinal correlation. Note that if the covariates, such as the education level, in the health care utilization data, collected from the same individual over a period of time, are time dependent, then the longitudinal lag correlations for the same individual will be nonstationary. It is of main interest to find β, the effects of the covariates on the count responses of an individual after taking the nonstationary familial and longitudinal correlations into account.

10.1.1 Marginal Mean and Variance

10.1.1.1 Conditional Marginal Mean and Variance

Suppose that conditional on the random family effect γ_i, y_{ijt} follows the Poisson density given by

$$f(y_{ijt}|\gamma_i) = \frac{1}{y_{ijt}!}\exp[y_{ijt}\log(\mu^*_{ijt}) - \mu^*_{ijt}], \tag{10.1}$$

where

$$\mu^*_{ijt} = \exp(x'_{ijt}\beta + \gamma_i).$$

We denote the marginal distribution in (10.1) by

$$y_{ijt}|\gamma_i \sim \text{Poi}(\mu^*_{ijt}).$$

This distribution yields the marginal mean and the variance of y_{ijt} conditional on γ_i as

$$E(Y_{ijt}|\gamma_i) = \text{var}(Y_{ijt}|\gamma_i) = \mu^*_{ijt}. \tag{10.2}$$

10.1.1.1 Unconditional Marginal Mean and Variance

As in Chapter 4, we assume that $\gamma_i \sim N(0, \sigma^2_\gamma)$ [Breslow and Clayton (1993); Jiang (1998); Sutradhar (2004)]. It then follows that the unconditional mean and the variance of y_{ijt} are given by

$$\begin{aligned} E[Y_{ijt}] &= \mu_{ijt} = \exp(x'_{ijt}\beta + \frac{\sigma^2_\gamma}{2}) \\ \text{var}[Y_{ijt}] &= \mu_{ijt} + \{\exp(\sigma^2_\gamma) - 1\}\mu^2_{ijt}. \end{aligned} \tag{10.3}$$

10.1.2 Nonstationary Autocorrelation Models

As an extension of the longitudinal models discussed in Chapters 6 and 7, Sutradhar (2010) has introduced a class of autocorrelation models for familial longitudinal data. These models are referred to as the generalized linear longitudinal mixed models (GLLMMs), appropriate for both count and binary data. In this section, we consider the GLLMMs from Sutradhar (2010) for the count data.

10.1.2.1 Conditional AR(1) Model

Initially consider the nonstationary AR(1) longitudinal model (6.44) for repeated count data for an individual, discussed in Section 6.5.1.1. We now assume that conditional on the random family effects γ_i, the AR(1) model (6.44) is appropriate for the jth ($j = 1,\ldots,n_i$) member of the ith family. Thus, we write

$$y_{ijt}|\gamma_i = \rho * [y_{ij,t-1}|\gamma_i] + d_{ijt}|\gamma_i, \text{ for } t = 2,\ldots,T, \tag{10.4}$$

where

$$\rho * y_{ij,t-1} = \sum_{s=1}^{y_{ij,t-1}} b_s(\rho),$$

with $Pr[b_s(\rho) = 1] = \rho$ and $Pr[b_s(\rho) = 0] = 1-\rho$, but,

$$d_{ijt}|\gamma_i \sim Poi(\mu^*_{ijt} - \rho\mu^*_{ij,t-1}),$$

with $\mu^*_{ijt} = \exp(x'_{ijt}\beta + \gamma_i)$, as in (10.2) for all $t = 1,\ldots,T$. Also, suppose that conditional on γ_i, d_{ijt} is independent of $z_{ij,t-1} = \rho * y_{ij,t-1}$. Now by assuming that for $t = 1$,

$$y_{ij1}|\gamma_i \sim \text{Poi}(\mu^*_{ij1}),$$

in the fashion similar to that of (6.45) – (6.46), it can be shown that the model (10.4) produces the marginal mean and the variance given by

$$E(Y_{ijt}|\gamma_i) = \text{var}(Y_{ijt}|\gamma_i) = \mu^*_{ijt}, \tag{10.5}$$

the same as in (10.2), and it also produces the autocorrelations

$$\text{corr}(Y_{iju}, Y_{ijt}|\gamma_i) = \rho^{|t-u|}\left[\frac{\mu^*_{iju}}{\mu^*_{ijt}}\right]^{1/2} = \rho^{|t-u|}\left[\frac{\mu_{iju}}{\mu_{ijt}}\right]^{1/2} = \rho^{|t-u|} r_{ijut}, \tag{10.6}$$

with $r_{ijut} = \exp\{-\frac{1}{2}(x_{ijt} - x_{iju})'\beta\}$, and μ_{ijt} has the formula as in (10.3). As far as the longitudinal correlations between two members of a family are concerned, we assume that at any two time points, the responses of any two members are conditionally independent. In notation,

$$\text{cov}[\{Y_{iju}, Y_{ikt}\}|\gamma_i] = 0, \text{ for } j \neq k. \tag{10.7}$$

10.1.2.1.1. Unconditional Mean, Variance, and Correlation Structure

Based on the assumption that $\gamma_i \overset{iid}{\sim} N(0, \sigma_\gamma^2)$, it follows from $(10.5)-(10.7)$ that the unconditional mean, variance and the covariances have the formulas given by

$$E(Y_{ijt}) = \mu_{ijt} = \exp(x'_{ijt}\beta + \sigma_\gamma^2/2), \text{for all } i, j, \text{ and } t$$
$$\text{cov}(Y_{iju}, Y_{ikt}) = \begin{cases} \mu_{ijt} + [\exp(\sigma_\gamma^2) - 1]\mu_{ijt}^2 & \text{for } k = j; u = t \\ \rho^{t-u}\mu_{iju} + [\exp(\sigma_\gamma^2) - 1]\mu_{iju}\mu_{ijt} & \text{for } k = j; u < t \\ [\exp(\sigma_\gamma^2) - 1]\mu_{iju}\mu_{ikt} & \text{for } k \neq j; u \leq t. \end{cases} \tag{10.8}$$

10.1.2.2 Conditional MA(1) Model

Similar to the AR(1) case, the extended familial longitudinal MA(1) model as an extension of (6.49), may be written as

$$\begin{aligned} y_{ij1}|\gamma_i &\sim \text{Poi}(\mu_{ij1}^* = \exp(x'_{ij1}\beta + \gamma_i)) \\ y_{ijt}|\gamma_i &= \rho * [d_{ij,t-1}|\gamma_i] + d_{ijt}|\gamma_i, \text{ for } j = 1, \ldots, n_i;\ t = 2, \ldots, T, \end{aligned} \tag{10.9}$$

where the distributional assumptions, conditional on γ_i remain the same as those for the longitudinal MA(1) model (6.49). That is,

$$d_{ijt}|\gamma_i \overset{iid}{\sim} Poi\left[\sum_{k=0}^{t-1} (-\rho)^k \mu_{ij,t-k}^*\right] \text{ for all } t = 1, \ldots, T.$$

Also, it is assumed that

$$\text{cov}[\{Y_{iju}, Y_{ikt}\}|\gamma_i] = 0, \text{ for } j \neq k. \tag{10.10}$$

This conditional MA(1) model (10.9) produces the same conditional mean and the variance [given by (10.5)] as in the AR(1) case. The conditional correlation is, however, different and is given by

$$\begin{aligned} c_{ij,ut}^{(ns)}(x_{iju}, x_{ijt}, \rho, \gamma_i) &= \text{corr}[(Y_{iju}, Y_{ijt})|\gamma_i] \\ &= \begin{cases} \dfrac{\rho\{\Sigma_{k=0}^{\min(u,t)-1}(-\rho)^k \mu_{ij,\min(u,t)-k}^*\}}{\sqrt{\mu_{iju}^* \mu_{ijt}^*}} & \text{for } |u-t| = 1 \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{10.11}$$

which under the stationary case reduces to

$$c_{ij,ut}^*(\rho) = \text{corr}[(Y_{iju}, Y_{ijt})|\gamma_i] = \begin{cases} \rho\{\Sigma_{k=0}^{\infty}(-\rho)^k = \frac{\rho}{1+\rho} & \text{for} |u-t| = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{10.12}$$

the same as in the longitudinal case (6.52), as expected.

10.1.2.2.1. Unconditional Mean, Variance, and Correlation Structure

Based on the assumption that $\gamma_i \stackrel{iid}{\sim} N(0, \sigma_\gamma^2)$, it follows from (10.9) – (10.11) that the unconditional mean and the variance have the formulas as in (10.8) under the AR(1) case, but, the unconditional covariances have the formulas

$$\text{cov}(Y_{iju}, Y_{ikt}) = \begin{cases} \rho\{\sum_{k=0}^{\min(u,t)-1} (-\rho)^k \mu_{ij,\min(u,t)-k}\} \\ +[\exp(\sigma_\gamma^2) - 1]\mu_{iju}\mu_{ikt} & \text{for } k = j; |u-t| = 1 \\ [\exp(\sigma_\gamma^2) - 1]\mu_{iju}\mu_{ikt} & \text{for } k = j; |t-u| > 1 \\ [\exp(\sigma_\gamma^2) - 1]\mu_{iju}\mu_{ikt} & \text{for } k \neq j; u \leq t. \end{cases} \tag{10.13}$$

10.1.2.3 An Alternative Conditional MA(1) Model

The MA(1) model in Section 10.1.2.2 produces the same mean and variance as those by the AR(1) model discussed in Section 10.1.2.1. However, as introduced in Chapter 6 (see Section 6.6), one may use a certain alternative MA(1) model which produces Gaussian MA(1) correlations in the stationary case, but the mean and the variance under the alternative model can be different from those under the AR(1) model. Following Sutradhar, Jowaheer, and Sneddon (2008), we consider such an MA(1) model given by

$$y_{ijt}|\gamma_i = \rho \circ [d_{ij,t-1}|\gamma_i] + d_{ijt}|\gamma_i. \tag{10.14}$$

Suppose that conditional on γ_i, d_{ijt}, and $d_{ij,t-1}$ follow the Poisson distributions given by

$$d_{ijt}|\gamma_i \sim \text{Poi}(\mu^*_{ijt}/(1+\rho)), \text{ and } d_{ij,t-1}|\gamma_i \sim \text{Poi}(\mu^*_{ij,t-1}/(1+\rho)), \tag{10.15}$$

respectively, with $\mu^*_{ijt} = \exp(x'_{ijt}\beta + \gamma_i)$. The model in (10.14) – (10.15) produces the conditional mean, variance, and conditional correlations as

$$E[Y_{ijt}|\gamma_i] = \text{var}[Y_{ijt}|\gamma_i] = \frac{\mu^*_{ijt} + \rho\mu^*_{ij,t-1}}{1+\rho} \tag{10.16}$$

and

$$\text{corr}[(Y_{iju}, Y_{ikt})|\gamma_i] = \begin{cases} \frac{\rho}{1+\rho}\left[\frac{\mu^*_{iju}}{\mu^*_{ijt}}\right]^{1/2} & \text{for } |t-u| = 1 \\ 0 & \text{for } \ell > 1. \end{cases} \tag{10.17}$$

It is clear from (10.16) that the conditional mean and the variance are different from those of the MA(1) model considered in the last section [see (10.5) for mean and variance]. The formula for the conditional autocorrelations given in (10.17) is, however, much simpler than that of (10.11). Note that because the autocorrelations under the current as well as the previous MA(1) models are functions of the time dependent covariates, using the stationary MA(1) model based 'working' autocorrelations with $\rho_1 = \rho/(1+\rho)$, and $\rho_\ell = 0$, for $\ell = 2, 3, \ldots$ in place of the conditional correlations (10.17) or (10.11), may lead to an inconsistent estimate for the correlation parameter.

10.1.2.3.1 Unconditional First− and Second-Order Moments

By using the assumption that $\gamma_i \overset{iid}{\sim} N(0, \sigma_\gamma^2)$, it follows from (10.16) that the unconditional first moment under the current MA(1) model is given by

$$E(Y_{ijt}) = [\mu_{ijt} + \rho\mu_{ij,t-1}]/(1+\rho) = \tilde{\mu}_{ijt} \text{ for } j = 1, \ldots, n_i;\ t = 1, \ldots, T, \quad (10.18)$$

where $\mu_{ijt} = \exp(x'_{ijt}\beta + \sigma_\gamma^2/2)$. Furthermore, for $u \leq t$, the unconditional variances and covariances have the formulas

$$\text{cov}(Y_{iju}, Y_{ikt}) = \begin{cases} \tilde{\mu}_{ijt} + [\exp(\sigma_\gamma^2) - 1]\tilde{\mu}_{ijt}^2 & \text{for } k = j; u = t \\ \rho\mu_{iju}/(1+\rho) + [\exp(\sigma_\gamma^2) - 1]\tilde{\mu}_{ijt}\tilde{\mu}_{iju} & \text{for } k = j; t - u = 1 \\ [\exp(\sigma_\gamma^2) - 1]\tilde{\mu}_{iju}\tilde{\mu}_{ijt} & \text{for } k = j; t - u > 1 \\ [\exp(\sigma_\gamma^2) - 1]\tilde{\mu}_{iju}\tilde{\mu}_{ikt} & \text{for } k \neq j; u \leq t. \end{cases} \quad (10.19)$$

It is clear from (10.19) that the unconditional covariances are functions of the time-dependent covariates in a complicated way, which is an effect of the involvement of the time dependent covariates in the conditional correlations.

10.1.2.4 Conditional EQC Model

The EQC model in the familial longitudinal setup may be written by extending the longitudinal EQC model given in (6.53). This extended EQC model is given by

$$\begin{aligned} y_{ij1}|\gamma_i &\sim Poi(\mu_{ij1}^* = \exp(x'_{ij1}\beta + \gamma_i)) \\ y_{ijt}|\gamma_i &= \rho * [y_{ij1}|\gamma_i] + d_{ijt}|\gamma_i, \text{ for } j = 1, \ldots, n_i;\ t = 2, \ldots, T, \end{aligned} \quad (10.20)$$

Also it is assumed that d_{ijt} for $t = 2, \ldots, T$, are independent of y_{ij1}. It then follows that the conditional mean and the variance are given by

$$E[Y_{ijt}|\gamma_i] = \text{var}[Y_{ijt}|\gamma_i] = \mu_{ijt}^* = \exp(x'_{ijt}\beta + \gamma_i),$$

the same as those under the AR(1) model (10.4) and MA(1) model (10.9), for all $t = 1,\ldots,T$. Furthermore, for $u < t$, the conditional autocovariances for the jth member of the ith family, are given by

$$\text{cov}[(Y_{iju}, Y_{ijt})|\gamma_i] = \rho\mu^*_{ij1}, \tag{10.21}$$

yielding the nonstationary conditional correlation structure

$$c^{(ns)}_{i,ut}(x_{iju}, x_{ijt}, \rho, \gamma_i) = \text{corr}[(Y_{iju}, Y_{ijt})|\gamma_i] = \frac{\rho\mu^*_{ij1}}{\sqrt{\mu^*_{iju}\mu^*_{ijt}}}, \tag{10.22}$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \frac{\mu^*_{ijt}}{\mu^*_{ij1}}\right], \quad t = 2,\ldots,T.$$

For two different members, that is, for $j \neq k$, similar to AR(1) and MA(1) models, we assume that

$$\text{cov}[(Y_{iju}, Y_{ikt})|\gamma_i] = 0. \tag{10.23}$$

10.1.2.4.1. Unconditional Mean, Variance, and Correlation Structure

Based on the assumption that $\gamma_i \overset{iid}{\sim} N(0, \sigma^2_\gamma)$, it follows from $(10.20) - (10.21)$ that the unconditional mean, variance, and the covariances have the formulas given by

$$E(Y_{ijt}) = \mu_{ijt} = \exp(x'_{ijt}\beta + \sigma^2_\gamma/2), \text{ for all } i, j, \text{ and } t$$

$$\text{cov}(Y_{iju}, Y_{ikt}) = \begin{cases} \mu_{ijt} + [\exp(\sigma^2_\gamma) - 1]\mu^2_{ijt} & \text{for } k = j; u = t \\ \rho\mu_{ij1} + [\exp(\sigma^2_\gamma) - 1]\mu_{iju}\mu_{ijt} & \text{for } k = j; u < t \\ [\exp(\sigma^2_\gamma) - 1]\mu_{iju}\mu_{ikt} & \text{for } k \neq j; u \leq t. \end{cases} \tag{10.24}$$

Note that the unconditional means and the variances in (10.24) are the same as in (10.8) under the AR(1) process. They are also the same as those under the conditional MA(1) model in (10.9), but different from those under the conditional MA(1) model given by (10.14). As far as the unconditional covariances (10.24) under the EQC model are concerned, they are generally different from those under the AR(1) and MA(1) models, especially for the same member, that is, when $j = k$.

10.2 Parameter Estimation

In the familial longitudinal setup, we need to estimate the regression effects β, the random effects variance σ^2_γ (also referred to as the familial correlation index parameter), and the longitudinal correlation index parameter ρ. We assume that the correlation model for the data is known or identified in the fashion similar to that of Section 6.5.3, where we dealt with longitudinal correlation model selection. For convenience, we now discuss the estimation under one model, namely under the conditional AR(1) model (10.4) – (10.7). For the purpose, we solve the appropriate GQL estimating equations to estimate β and σ^2_γ, and use the MM (method of moments) to estimate ρ.

10.2.1 Estimation of Parameters Under Conditional AR(1) Model

10.2.1.1 GQL Estimation of Regression Parameter β

Let $y_{ij} = (y_{ij1}, \ldots, y_{ijt}, \ldots, y_{ijT})'$ denote the $T \times 1$ repeated responses recorded over T occasions for the jth $(j = 1, \ldots, n_i)$ member of the ith $(i = 1, \ldots, K)$ family. Furthermore, let

$$y_i = (y'_{i1}, \ldots, y'_{ij}, \ldots, y'_{in_i})'$$

denote the $n_iT \times 1$ vector of count responses for the ith family.

We now write the $n_iT \times 1$ unconditional mean vector of y_i as

$$\mu_i(\beta, \sigma^2_\gamma) = (\mu'_{i1}(\beta, \sigma^2_\gamma), \ldots, \mu'_{ij}(\beta, \sigma^2_\gamma), \ldots, \mu'_{in_i}(\beta, \sigma^2_\gamma))', \tag{10.25}$$

where

$$\mu_{ij}(\beta, \sigma^2_\gamma) = (\mu_{ij1}(\beta, \sigma^2_\gamma), \ldots, \mu_{ijt}(\beta, \sigma^2_\gamma), \ldots, \mu_{ijT}(\beta, \sigma^2_\gamma))'$$

is the $T \times 1$ vector with μ_{ijt} as its general element. The formula for this general element is given by (10.8).

Also let the covariance elements defined by (10.8) constitute the T-dimensional diagonal matrices $\Sigma_{ijj}(\beta, \sigma^2_\gamma, \rho)$ for all $j = 1, \ldots, n_i$, and the off-diagonal matrices $\Sigma_{ijk}(\beta, \sigma^2_\gamma)$ for all $j \neq k$, $j,k = 1, \ldots, n_i$. It then follows that the $n_iT \times n_iT$ unconditional variance–covariance matrix of y_i can be expressed as

$$\Sigma_i(\beta, \sigma^2_\gamma, \rho) = \begin{pmatrix} \Sigma_{i11} & \Sigma_{i12} & \cdots & \Sigma_{i1k} & \cdots & \Sigma_{i1n_i} \\ & \Sigma_{i22} & \cdots & \Sigma_{i2k} & \cdots & \Sigma_{i2n_i} \\ & & & \vdots & & \\ & & & \Sigma_{ikk} & \cdots & \Sigma_{ikn_i} \\ & & & & \ddots & \vdots \\ & & & & & \Sigma_{in_in_i} \end{pmatrix}. \tag{10.26}$$

For known σ^2_γ and ρ, similar to the GQL estimation in Section 6.5.2, the GQL estimating equation for β in the present familial longitudinal setup, may be written as

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1}(y_i - \mu_i) = 0, \tag{10.27}$$

where μ_i and Σ_i are defined as in (10.25) and (10.26) respectively, and $\partial\mu_i/\partial\beta'$ is the $n_iT \times p$ first-order derivative matrix. Note that because

$$\mu_{ij} = (\mu_{ij1}, \ldots, \mu_{ijt}, \ldots, \mu_{ijT})'$$

with μ_{ijt} defined as in (10.8), the derivative of μ_i with respect to β' requires the derivation of μ_{ijt} with respect to β. To be specific, $\partial\mu_{ijt}/\partial\beta$ is the $p \times 1$ vector given as

$$\frac{\partial \mu_{ijt}}{\partial \beta} = \mu_{ijt} x_{ijt}, \tag{10.28}$$

where x_{ijt} is the $p \times 1$ vector of all covariates for the jth individual under the ith family at time t. Let $\hat{\beta}_{GQL}$ denote the GQL estimator of β, obtained by solving the estimating equation (10.27). This estimator is consistent, and it is highly efficient as the GQL estimating equation is unbiased as well as the weight matrix Σ_i is the true covariance matrix of y_i. Furthermore, by using the multivariate central limit theorem [see Mardia, Kent, and Bibby (1979, p.51), for example], it may be shown that $K^{\frac{1}{2}}(\hat{\beta}_{GQL} - \beta)$ has an asymptotic normal distribution, as $K \to \infty$, with mean zero and with covariance matrix given by

$$K \left(\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1} \frac{\partial \mu_i}{\partial \beta'} \right)^{-1}. \tag{10.29}$$

Remark that the computation of the estimate of β by (10.27) requires the estimates for σ_γ^2 and ρ. These parameters are estimated in the following two sections.

10.2.1.2 GQL Estimation of Familial Correlation Index Parameter σ_γ^2

In Chapter 8 (see Section 8.2.2.2), this variance component parameter was estimated in a longitudinal mixed model setup, where in addition to a time factor, the repeated count responses of an individual were also influenced by the individual's random effect. In the familial longitudinal setup, repeated responses of an individual family member are affected by a time factor and the random family effect. Consequently, we can generalize the GQL estimation of σ_γ^2 given in Section 8.2.2.2 to the familial longitudinal setup, as follows.

Let

$$u_{ij} = [u_{ij(s)}', u_{ij(p)}']' \tag{10.30}$$

be the $T(T+1)/2$-dimensional combined vector of squares and pairwise products for the jth $(j = 1, \ldots, n_i)$ member of the ith $(i = 1, \ldots, K)$ family, where

$$u_{ij(s)} = [y_{ij1}^2, \ldots, y_{ijt}^2, \ldots, y_{ijT}^2]' : T \times 1$$

$$u_{ij(p)} = [y_{ij1}y_{ij2}, \ldots, y_{ijt}y_{ijv}, \ldots, y_{ij(T-1)}y_{ijT}]' : \frac{T(T-1)}{2} \times 1.$$

Next, we write the $n_i T(T+1)/2$-dimensional vector of squares and distinct products for all n_i individuals in the ith family. Let u_i denote this vector and λ_i be its mean. That is,

$$\begin{aligned} u_i &= [u'_{i1}, \ldots, u'_{ij}, \ldots, u'_{in_i}]' \\ \lambda_i &= [\lambda'_{i1}, \ldots, \lambda'_{ij}, \ldots, \lambda'_{in_i}]', \end{aligned} \tag{10.31}$$

where

$$\lambda_{ij} = [\lambda'_{ij(s)}, \lambda'_{ij(p)}]',$$

with

$$\begin{aligned} \lambda_{ij(s)} &= [E(Y_{ij1}^2), \ldots, E(Y_{ijt}^2), \ldots, E(Y_{ijT}^2)]' \\ &= [\lambda_{ij,11}, \ldots, \lambda_{ij,tt}, \ldots, \lambda_{ij,TT}]' \end{aligned} \tag{10.32}$$

$$\begin{aligned} \lambda_{ij(p)} &= [E(Y_{ij1}Y_{ij2}), \ldots, E(Y_{ijv}Y_{ijt}), \ldots, E(Y_{ij(T-1)}Y_{ijT})]', \\ &= [\lambda_{ij,12}, \ldots, \lambda_{ij,vt}, \ldots, \lambda_{ij,T-1,T}]', \end{aligned} \tag{10.33}$$

where

$$\lambda_{ij,tt} = \mu_{ijt} + [\exp(\sigma_\gamma^2)]\mu_{ijt}^2$$

and for $v < t$

$$\lambda_{ij,vt} = \rho^{t-v}\mu_{ijv} + [\exp(\sigma_\gamma^2)]\mu_{ijv}\mu_{ijt}$$

by (10.8), with $\mu_{ijt} = \exp(x'_{ijt}\beta + \sigma_\gamma^2/2)$. Similar to (8.42), we may write the GQL estimating equation for σ_γ^2 as

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta, \sigma_\gamma^2, \rho)}{\partial \sigma_\gamma^2} \Omega_i^{-1}(\beta, \sigma_\gamma^2, \rho)[u_i - \lambda_i(\beta, \sigma_\gamma^2, \rho)] = 0, \tag{10.34}$$

where, unlike (8.42), $\Omega_i = \text{cov}[U_i]$ is the $\{n_i T(T+1)/2\} \times \{n_i T(T+1)/2\}$ covariance matrix of u_i, and $\partial\lambda_i'/\partial\sigma_\gamma^2$ is the $1 \times \{n_i T(T+1)/2\}$ vector of first derivative of λ_i with respect to σ_γ^2. The formulas for the elements of this derivative vector are available in Exercise 10.1. As far as the construction of Ω_i is concerned, one may use either $\Omega_i(\beta, \sigma_\gamma^2, \rho = 0)$ or the normality based approximation $\Omega_{iN}(\beta, \sigma_\gamma^2, \rho)$. However, following the notation for u_i from (10.31), we first express the Ω_i in (10.34) as

$$\Omega_i = \begin{bmatrix} \Omega_{i11} & \Omega_{i12} & \cdots & \Omega_{i1k} & \cdots & \Omega_{i1n_i} \\ & \Omega_{i22} & \cdots & \Omega_{i2k} & \cdots & \Omega_{i2n_i} \\ & & & \vdots & & \vdots \\ & & & \Omega_{ikk} & \cdots & \Omega_{ikn_i} \\ & & & & & \vdots \\ & & & & & \Omega_{in_in_i} \end{bmatrix}, \tag{10.35}$$

where $\Omega_{ijj} = \text{cov}(U_{ij})$ and $\Omega_{ijk} = \text{cov}(U_{ij}, U_{ik})$, for $j \neq k,\ j,k = 1,\ldots,n_i$. Note that the Ω_i matrix in (10.35) appears to be quite similar to the Σ_i matrix defined in (10.26). They are, however, different matrices. In (10.26) Σ_{ijk} is the $T \times T$ covariance matrix of $y_{ij} = (y_{ij1},\ldots,y_{ijT})'$ and $y_{ik} = (y_{ik1},\ldots,y_{ikT})'$, whereas Ω_{ijk} in (10.35) is the $\{T(T+1)/2\} \times \{T(T+1)/2\}$ covariance matrix of $u_{ij} = (u'_{ij(s)}, u'_{ij(p)})'$ and $u_{ik} = (u'_{ik(s)}, u'_{ik(p)})'$. Note that in order to construct the Ω_i in (10.35) it is sufficient to construct two matrices, namely, Ω_{ijj} and Ω_{ijk}, where Ω_{ijj} is the jth block diagonal matrix and Ω_{ijk} is the block off-diagonal matrix corresponding to the individuals j and k.

10.2.1.2.1 GQL(I) Estimation of σ^2_γ

When the Ω_i matrix in (10.34) is approximated based on independence (I) assumption, one writes the GQL(I) estimating equation for σ^2_γ as

$$\sum_{i=1}^{K} \frac{\partial \lambda_i'(\beta, \sigma^2_\gamma, \rho)}{\partial \sigma^2_\gamma} \Omega_i^{-1}(\beta, \sigma^2_\gamma, \rho = 0)[u_i - \lambda_i(\beta, \sigma^2_\gamma, \rho)] = 0, \tag{10.36}$$

where $\Omega_i(\beta, \sigma^2_\gamma, \rho = 0)$ is constructed as follows.

Construction of $\Omega_i(\beta, \sigma^2_\gamma, \rho = 0) \equiv \Omega_i(I)$

As mentioned in the last section, to construct the $\Omega_i(I)$ matrix as a substitute for the Ω_i matrix defined in (10.34), it is sufficient to compute $\Omega_{ijj}(I)$ for all $j = 1,\ldots,n_i$, and $\Omega_{ijk}(I)$ for all $j \neq k, j,k = 1,\ldots,n_i$. Note that when it is pretended that $\rho = 0$, it follows from (10.6) and (10.7) that for $v \neq t$,

$$\text{corr}\{(y_{ijv}, y_{ikt})|\gamma_i\} = 0, \ \text{ for all } j = k;\ j \neq k. \tag{10.37}$$

Thus to construct $\Omega_i(I)$, the submatrices $\Omega_{ijj}(I)$ and $\Omega_{ijk}(I)$ are constructed by using the conditional independence assumption in (10.37). For convenience, we write

$$\Omega_{ijj} = \begin{bmatrix} \text{cov}(U_{ij(s)}) & \text{var}(U_{ij(s)}, U'_{ij(p)}) \\ & \text{var}(U_{ij(p)}) \end{bmatrix}, \tag{10.38}$$

and

$$\Omega_{ijk} = \begin{bmatrix} \mathrm{cov}(U_{ij(s)}, U'_{ik(s)}) & \mathrm{cov}(U_{ij(s)}, U'_{ik(p)}) \\ & \mathrm{cov}(U_{ij(p)}, U'_{ik(p)}) \end{bmatrix}. \quad (10.39)$$

Now for the computation of the elements of $\Omega_{ijj}(I)$ and $\Omega_{ijk}(I)$ matrices, we use the conditional independence assumption (10.37) to derive the elements of these two matrices given in (10.38) and (10.39).

Note that for the computation of the elements of the $\Omega_{ijj}(I)$ in (10.38), it is sufficient to compute the formulas for

$$(a): \ (i) \ \ \mathrm{var}[Y_{ijt}^2], \ (ii)\ \mathrm{cov}[Y_{ijt}^2, Y_{ij\ell}^2], \ (iii)\ \mathrm{cov}[Y_{ijt}^2, Y_{ij\ell}Y_{iju}], \ (iv)\ \mathrm{var}[Y_{ijt}Y_{iju}], \text{ and } \ (v)\ \ \mathrm{cov}[Y_{ijt}Y_{iju}, Y_{ij\ell}Y_{ijv}]. \quad (10.40)$$

Similarly, for the computation of the elements of the $\Omega_{ijk}(I)$ in (10.39), it is sufficient to compute the formulas for

$$(b): \ (i)\ \mathrm{cov}[Y_{iju}^2, Y_{ikt}^2], \ (ii)\ \mathrm{cov}[Y_{ijt}^2, Y_{ik\ell}Y_{iku}], \ (iii)\ \mathrm{cov}[Y_{ijt}Y_{iju}, Y_{ik\ell}Y_{ikv}]. \quad (10.41)$$

Note that the computation of the moments in (10.40) and (10.41) requires the formulas for various unconditional fourth− as well as unconditional second-order moments. The formulas for the unconditional second-order moments, namely

$$\lambda_{ij,tt} = E[Y_{ijt}^2] \text{ and } \lambda_{ij,vt} = E[Y_{ijv}Y_{ijt}]$$

are already given in (10.32) and (10.33). Next, let

$$\phi^*_{i,jk,tu\ell v} = E[\{Y_{ijt}Y_{iju}Y_{ik\ell}Y_{ikv}\}|\gamma_i, \rho = 0] \ \text{ and } \ \phi_{i,jk,tu\ell v} = E[\{Y_{ijt}Y_{iju}Y_{ik\ell}Y_{ikv}\}|\rho = 0], \quad (10.42)$$

denote the conditional (on γ_i) and corresponding unconditional fourth-order moments evaluated at $\rho = 0$, for any $j,k = 1,\ldots,n_i$, and at times $t,u = 1,\ldots,T$, for the jth member and at times $\ell, v = 1,\ldots,T$, for the kth member. The formulas for these conditional and unconditional moments evaluated at $\rho = 0$, are available in Exercises 10.2 and 10.3, which may be used to compute the covariances in (10.40) and (10.41). For example, $\mathrm{cov}[Y_{ijt}Y_{iju}, Y_{ik\ell}Y_{ikv}]$ is calculated by using the formula

$$\mathrm{cov}[Y_{ijt}Y_{iju}, Y_{ik\ell}Y_{ikv}] = [\phi_{i,jk,tu\ell v} - \lambda_{ij,tu}\lambda_{ik,\ell v}].$$

Note that similar to that of Section 8.2.2.3.2, the asymptotic variance of the GQL estimator of σ_γ^2 obtained from (10.36) has the formula

$$\mathrm{var}(\hat{\sigma}^2_{GQL}) = Lt_{K\to\infty} \left[\sum_{i=1}^K \frac{\partial \lambda_i'}{\partial \sigma_\gamma^2} \Omega_i^{-1}(\beta, \sigma_\gamma^2, \rho = 0) \frac{\partial \lambda_i'}{\partial \sigma_\gamma^2} \right]^{-2}$$
$$\times \left[\sum_{i=1}^K \frac{\partial \lambda_i'}{\partial \sigma_\gamma^2} \Omega_i^{-1}(\beta, \sigma_\gamma^2, \rho = 0) \Omega_i(\beta, \sigma_\gamma^2, \rho) \right.$$

$$\times \Omega_i^{-1}(\beta,\sigma_\gamma^2,\rho=0)\frac{\partial \lambda_i'}{\partial \sigma_\gamma^2}\Bigg], \tag{10.43}$$

which may be consistently estimated by

$$\begin{aligned}\hat{\text{var}}(\hat{\sigma}^2_{GQL}) = &\left[\sum_{i=1}^{K}\frac{\partial \lambda_i'}{\partial \sigma_\gamma^2}\Omega_i^{-1}(\beta,\sigma_\gamma^2,\rho=0)\frac{\partial \lambda_i'}{\partial \sigma_\gamma^2}\right]^{-2}\\ &\times\left[\sum_{i=1}^{K}\frac{\partial \lambda_i'}{\partial \sigma_\gamma^2}\Omega_i^{-1}(\beta,\sigma_\gamma^2,\rho=0)(u_i-\hat{\lambda}_i)(u_i-\hat{\lambda}_i)'\right.\\ &\left.\times \Omega_i^{-1}(\beta,\sigma_\gamma^2,\rho=0)\frac{\partial \lambda_i}{\partial \sigma_\gamma^2}\right],\end{aligned} \tag{10.44}$$

where $\hat{\lambda}_i$ is computed by using $\hat{\beta}_{GQL}$ and $\hat{\sigma}^2_{\gamma,GQL}$ in the formula for λ_i given by (10.31).

10.2.1.2.2 GQL(N) Estimation of σ_γ^2

When the Ω_i matrix in (10.34) is approximated by pretending that the data follow a normal (N) distribution, one writes the GQL(N) estimating equation for σ_γ^2 as

$$\sum_{i=1}^{K}\frac{\partial \lambda_i'(\beta,\sigma_\gamma^2,\rho)}{\partial \sigma_\gamma^2}\Omega_{iN}^{-1}(\beta,\sigma_\gamma^2,\rho)[u_i-\lambda_i(\beta,\sigma_\gamma^2,\rho)] = 0, \tag{10.45}$$

where $\Omega_{iN}(\beta,\sigma_\gamma^2,\rho)$ is constructed as follows.

Construction of $\Omega_{iN}(\beta,\sigma_\gamma^2,\rho)$

To construct this normality assumption based fourth-order moment matrix, it is sufficient to construct the normality (N) assumption based two general matrices, $\Omega_{ijj}(N)$ and $\Omega_{ijk}(N)$, where Ω_{ijj} and Ω_{ijk} are defined by (10.38) and (10.39), respectively. Note that to construct the $\Omega_{iN}(\cdot)$ matrix under the normality assumption, one pretends that

$$y_i = (y_{i1}',\ldots,y_{ij}',\ldots,y_{in_i}')' \; : n_iT\times 1$$

count response vector follows the n_iT-dimensional multivariate normal vector but with true Poisson mean vector μ_i (10.25) and Poisson AR(1) correlation structure based covariance matrix Σ_i (10.26). Express the Σ_i matrix as

$$\Sigma_i(\beta,\sigma_\gamma^2,\rho) = (\sigma_{i,jk,ut}) : \; n_iT\times n_iT, \tag{10.46}$$

where the formulas for

$$\sigma_{i,jk,ut} = \text{cov}[Y_{iju},Y_{ikt}]$$

for all $j,k = 1,\ldots,n_i$, and $u,t = 1,\ldots,T$, are given by (10.8).

Construction of $\Omega_{ijj}(N)$

Recall from (10.30) that Ω_{ijj} is the $\{T(T+1)/2\} \times \{T(T+1)/2\}$ covariance matrix of

$$u_{ij} = (u'_{ij(s)}, u'_{ij(p)})',$$

$u_{ij(s)}$ being the $T \times 1$ vector of squares of the elements of $y_{ij} = (y_{ij1},\ldots,y_{ijT})'$, and $u_{ij(p)}$ is the $\{T(T-1)/2 \times 1\}$ vector of distinct pairwise products of the elements of y_{ij}. Now by using the marginal property of the multivariate normal distribution, we write by following (10.46) that

$$y_{ij} = [y_{ij1},\ldots,y_{ijt},\ldots,y_{ijT}] \sim N_T(\mu_{ij}, \Sigma_{ijj}), \tag{10.47}$$

where

$$\Sigma_{ijj} = (\sigma_{i,jj,ut})$$

with $\sigma_{i,jj,ut} = \text{cov}[Y_{iju}, Y_{ijt}]$ as given in (10.8). Further note that under the normality assumption, one writes

$$E(Y_{ijt} - \mu_{ijt})(Y_{ijv} - \mu_{ijv})(Y_{ijr} - \mu_{ijr}) = 0,$$

and

$$\begin{aligned} & E(Y_{ijt} - \mu_{ijt})(Y_{ijv} - \mu_{ijv})(Y_{ijr} - \mu_{ijr})(Y_{ijd} - \mu_{ijd}) \\ &= \sigma_{i,jj,tv}\sigma_{i,jj,rd} + \sigma_{i,jj,tr}\sigma_{i,jj,vd} + \sigma_{i,jj,td}\sigma_{i,jj,vr}, \end{aligned} \tag{10.48}$$

yielding, by $(10.32)-(10.33)$,

$$\begin{aligned} \delta_{i,jjj,tvr} &= E[Y_{ijt}Y_{ijv}Y_{ijr}] \\ &= \lambda_{ij,tv}\mu_{ijr} + \lambda_{ij,tr}\mu_{ijv} + \lambda_{ij,vr}\mu_{ijt} - 2\mu_{ijt}\mu_{ijv}\mu_{ijr}, \end{aligned} \tag{10.49}$$

and

$$\begin{aligned} \phi_{i,jjjj,tvrd} &= E[Y_{ijt}Y_{ijv}Y_{ijr}Y_{ijd}] \\ &= [\sigma_{i,jj,tv}\sigma_{i,jj,rd} + \sigma_{i,jj,tr}\sigma_{i,jj,vd} + \sigma_{i,jj,td}\sigma_{i,jj,vr} \\ &\quad + \delta_{i,jjj,tvr}\mu_{ijd} + \delta_{i,jjj,tvd}\mu_{ijr} + \delta_{i,jjj,trd}\mu_{ijv} + \delta_{i,jjj,vrd}\mu_{ijt}] \\ &\quad - [\lambda_{ij,tv}\mu_{ijr}\mu_{ijd} + \lambda_{ij,tr}\mu_{ijv}\mu_{ijd} + \lambda_{ij,vr}\mu_{ijt}\mu_{ijd} \\ &\quad + \lambda_{ij,td}\mu_{ijv}\mu_{ijr} + \lambda_{ij,vd}\mu_{ijt}\mu_{ijr} + \lambda_{ij,rd}\mu_{ijt}\mu_{ijv}] \\ &\quad + 3\mu_{ijt}\mu_{ijv}\mu_{ijr}\mu_{ijd}. \end{aligned} \tag{10.50}$$

Next, by putting $v = t$ and $d = r$ in (10.50), one easily obtains $\phi_{i,jjjj,ttrr} = E(Y_{ijt}^2 Y_{ijr}^2)$ yielding the (t,r)th element of the $\mathrm{cov}(U_{ij(s)})$ matrix as

$$\mathrm{cov}(Y_{ijt}^2, Y_{ijr}^2) = \phi_{i,jjjj,ttrr} - \lambda_{ij,tt}\lambda_{ij,rr}, \tag{10.51}$$

where $\lambda_{ij,tt}$ is given by (10.32).

In the manner similar to that of (10.51), one may compute any elements of the $\mathrm{cov}(U_{ij(s)}, U_{ij(p)})$ and $\mathrm{var}(U_{ij(p)})$ matrices. For example, the covariance between the products $y_{ijt}y_{ijv}$ and $y_{ijr}y_{ijd}$ may be obtained as

$$\mathrm{cov}(Y_{ijt}Y_{ijv}, Y_{ijr}Y_{ijd}) = \phi_{i,jjjj,tvrd} - \lambda_{ij,tv}\lambda_{ij,rd}. \tag{10.52}$$

This completes the construction of the $\Omega_{ijj}(N)$ matrix for the Ω_{iN} matrix in (10.45).

Construction of $\Omega_{ijk}(N)$ Matrices for Cases when $j \neq k$

Recall from (10.39) that Ω_{ijk} is the $\{T(T+1)/2\} \times \{T(T+1)/2\}$ covariance matrix of $u_{ij} = (u'_{ij(s)}, u'_{ij(p)})'$ and $u_{ik} = (u'_{ik(s)}, u'_{ik(p)})'$ for $j \neq k,\ j,k = 1,\ldots,n_i$. Note that in order to obtain various fourth-order moments similar to (10.50) to construct this $\Omega_{ijk}(N)$ matrix for two selected members $j \neq k$, it is appropriate to construct a stacked random vector

$$y^*_{i,jk} = [y'_{ij}, y'_{ik}]'$$

which under the normality assumption follows the $T(T+1)$-dimensional normal vector with mean

$$\mu^*_{i,jk} = [\mu'_{ij}, \mu'_{ik}]'$$

and covariance matrix

$$\Sigma^*_{ijk} = \begin{bmatrix} \Sigma_{ijj} & \Sigma_{ijk} \\ & \Sigma_{ikk} \end{bmatrix} = (\sigma^*_{i,jk,ut}),\ T(T+1) \times T(T+1), \tag{10.53}$$

where

$$\begin{aligned} \sigma^*_{i,jk,ut} &= \mathrm{cov}[Y_{iju}, Y_{ikt}] \\ &= E[Y_{iju}Y_{ikt}] - \mu_{iju}\mu_{ikt} = \lambda_{i,jk,ut} - \mu_{iju}\mu_{ikt} \\ &= [\exp(\sigma^2_\gamma) - 1]\mu_{iju}\mu_{ikt}, \end{aligned} \tag{10.54}$$

by (10.8). By following (10.48) and (10.49), and by using the notation $\lambda_{i,jk,ut}$ from (10 .54), one may compute the third− and fourth-order raw moments as

$$\begin{aligned} \delta_{i,jjk,tvr} &= E[Y_{ijt}Y_{ijv}Y_{ikr}] \\ &= \lambda_{ij,tv}\mu_{ijr} + \lambda_{i,jk,tr}\mu_{ijv} + \lambda_{i,jk,vr}\mu_{ijt} - 2\mu_{ijt}\mu_{ijv}\mu_{ikr}, \end{aligned} \tag{10.55}$$

$$\begin{aligned} \delta_{i,jkk,tvr} &= E[Y_{ijt}Y_{ikv}Y_{ikr}] \\ &= \lambda_{i,jk,tv}\mu_{ijr} + \lambda_{i,jk,tr}\mu_{ijv} + \lambda_{ik,vr}\mu_{ijt} - 2\mu_{ijt}\mu_{ikv}\mu_{ikr}, \end{aligned} \tag{10.56}$$

and

$$\begin{aligned}
\phi_{i,jjkk,tvrd} &= E[Y_{ijt}Y_{ijv}Y_{ikr}Y_{ikd}] \\
&= [\sigma^*_{i,jj,tv}\sigma^*_{i,kk,rd} + \sigma^*_{i,jk,tr}\sigma^*_{i,jk,vd} + \sigma^*_{i,jk,td}\sigma^*_{i,jk,vr} \\
&\quad + \delta_{i,jjk,tvr}\mu_{ijd} + \delta_{i,jjk,tvd}\mu_{ijr} + \delta_{i,jkk,trd}\mu_{ijv} + \delta_{i,jkk,vrd}\mu_{ijt}] \\
&\quad - [\lambda_{ij,tv}\mu_{ijr}\mu_{ijd} + \lambda_{i,jk,tr}\mu_{ijv}\mu_{ikd} + \lambda_{i,jk,vr}\mu_{ijt}\mu_{ikd} \\
&\quad + \lambda_{i,jk,td}\mu_{ijv}\mu_{ikr} + \lambda_{i,jk,vd}\mu_{ijt}\mu_{ikr} + \lambda_{ik,rd}\mu_{ijt}\mu_{ijv}] \\
&\quad + 3\mu_{ijt}\mu_{ijv}\mu_{ikr}\mu_{ikd}. \qquad (10.57)
\end{aligned}$$

Now by using (10.57), one can compute any elements of the $\Omega_{ijk}(N)$ matrix. For example,

$$\text{cov}(Y^2_{ijt}, Y^2_{ikv}) = \phi_{i,jjkk,ttvv} - \lambda_{ij,tt}\lambda_{ik,vv} \qquad (10.58)$$

$$\text{cov}(Y^2_{ijt}, Y_{ikv}Y_{ikr}) = \phi_{i,jjkk,ttvr} - \lambda_{ij,tt}\lambda_{ik,vr} \qquad (10.59)$$

$$\text{cov}(Y_{ijt}Y_{ijv}, Y^2_{ikr}) = \phi_{i,jjkk,tvrr} - \lambda_{ij,tv}\lambda_{ik,rr} \qquad (10.60)$$

$$\text{cov}(Y_{ijt}Y_{ijv}, Y_{ikr}Y_{ikd}) = \phi_{i,jjkk,tvrd} - \lambda_{ij,tv}\lambda_{ik,rd}. \qquad (10.61)$$

This completes the construction of the $\Omega_{ijk}(N)$ matrices.

10.2.1.3 Estimation of Longitudinal Correlation Index Parameter ρ

Note that the iterative solution of the estimating equation (10.27) for β, and (10.36) or (10.45) for σ^2_γ requires a consistent estimator for the longitudinal correlation index parameter ρ. This consistent estimation for ρ may be achieved by using the method of moments. Recall from (10.8) that under the AR(1) process, the lag 1 covariance between y_{ijt} and $y_{ij,t+1}$ is given by

$$\text{cov}(Y_{ijt}, Y_{ij,t+1}) = \rho\mu_{ijt} + [\exp(\sigma^2_\gamma) - 1]\mu_{ijt}\mu_{ij,t+1}.$$

For known β and σ^2_γ, one may then obtain the moment estimator of ρ, which is consistent, by equating the sample lag 1 autocovariance with its population counterpart. To be specific, the moment estimator of ρ under the AR(1) process has the formula given by

$$\hat{\rho}_M = \frac{a_1 - b_1}{c_1}, \qquad (10.62)$$

where

$$a_1 = \frac{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-1} \tilde{y}_{ijt}\tilde{y}_{ij(t+1)} / \left\{(T-1)\sum_{i=1}^K n_i\right\}}{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T} \tilde{y}^2_{ijt} / \left\{T\sum_{i=1}^K n_i\right\}},$$

$$b_1 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{t=1}^{T-1} \bar{\phi}_{ijjt(t+1)}}{(T-1)\sum_{i=1}^{K} n_i},$$

and

$$c_1 = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{t=1}^{T-1} \left[\bar{\phi}_{ijjt(t+1)}/q_{ijt(t+1)}\right] / \left\{(T-1)\sum_{i=1}^{K} n_i\right\},$$

where

$$\tilde{y}_{ijt} = \left\{(y_{ijt} - \mu_{ijt})/\sigma_{i,jj,tt}^{1/2}\right\},\ q_{ijt(t+1)} = \left\{[\exp(\sigma_\gamma^2) - 1](\mu_{ij(t+1)})\right\},\ \text{and}\ \bar{\phi}_{ijtt(t+1)}$$

is given as

$$\bar{\phi}_{ijjt(t+1)} = \frac{[\exp(\sigma_\gamma^2) - 1]}{[\{[\exp(\sigma_\gamma^2) - 1] + 1/\mu_{ijt}\}\{[\exp(\sigma_\gamma^2) - 1] + 1/\mu_{ij(t+1)}\}]^{\frac{1}{2}}}.$$

This correlation estimate from (10.62) is used in (10.27) and (10.36) [or (10.45)] to obtain further improved estimates of β and σ_γ^2, respectively, which are in turn used in (10.62) to obtain further improved estimate of ρ. This cycle of iteration continues until convergence.

10.2.2 Performance of the GQL Approach: A Simulation Study

To examine the performance of the GQL approach in the familial longitudinal set up, Sutradhar, Jowaheer, and Sneddon (2008) conducted a simulation study, and it was shown that the GQL approach works very well in estimating the regression effects β and familial correlation index parameter σ_γ^2, along with good performance of the moment approach for the estimation of the longitudinal correlation index parameter ρ. We explain this simulation study as follows.

10.2.2.1 Simulation Study with $p = 1$ Covariate

For this single covariate case, suppose that $\beta = 1.0$. Consider $K = 100$ families each with $n_i = 2 (i = 1,\ldots,K)$ members. Also suppose that count responses were collected from each member for a period of $T = 4$ time points. The covariates were chosen as follows. For the first member, the time-dependent covariate was chosen to be

$$x_{i1t1} = \begin{cases} (t^2 - 2.5)/8 & \text{for } i = 1,\ldots,K/2;\ t = 1,\ldots,4 \\ t^2/8 & \text{for } i = K/2+1,\ldots,K;\ t = 1,\ldots,4, \end{cases}$$

whereas for the second member the time-dependent covariate was taken as

$$x_{i2t1} = \begin{cases} 0.1 + (t-1) \times 0.25 & \text{for } i = 1, \ldots, K/4;\ t = 1, \ldots, 4 \\ (1 + t + t^2)/12 & \text{for } i = K/4+1, \ldots, 3K/4;\ t = 1, \ldots, 4 \\ (t^2 - 2.5)/8 & \text{for } i = 3K/4+1, \ldots, K;\ t = 1, \ldots, 4. \end{cases}$$

Table 10.1 Simulated mean, simulated standard error, and relative bias of the GQL estimates for parameters of the nonstationary familial longitudinal model with single covariate for selected values of σ_γ^2 and ρ; $K = 100$; $T = 4$; $\beta = 1.0$; 500 simulations.

Variance Component (σ_γ^2)	Correlation Parameter(ρ)	Quantity	Estimates $\hat{\beta}$	$\hat{\sigma_\gamma^2}$	$\hat{\rho}$
0.50	0.25	SM	0.968	0.488	0.279
		SSE	0.077	0.169	0.264
		RB	42	7	11
	0.75	SM	0.980	0.494	0.715
		SSE	0.047	0.134	0.161
		RB	43	4	22
0.75	0.25	SM	0.923	0.730	0.272
		SSE	0.146	0.269	0.269
		RB	53	7	8
	0.75	SM	0.955	0.707	0.627
		SSE	0.101	0.218	0.275
		RB	45	20	45

Next, we choose two values for the σ_γ^2 parameter, namely, $\sigma_\gamma^2 \equiv 0.50,\ 0.75$, and to reflect small and large longitudinal correlations we have chosen $\rho \equiv 0.25,\ 0.75$. Note that when $\sigma_\gamma^2 = 0$, the longitudinal mixed model reduces to the longitudinal fixed model. Furthermore, it is clear from (10.8) that a small increase in the value of σ_γ^2 will cause a large change in overdispersion for the data. Thus, even though in theory σ_γ^2 can take any nonnegative values, it seems to be more practical to consider only moderately large values for σ_γ^2 in the simulation study such as $\sigma_\gamma^2 = 0.5,\ 0.75$. We remark here that some of the existing estimation methods such as the penalized quasi-likelihood (PQL) approach [Breslow and Lin (1995, p. 90)] cannot even unbiasedly estimate this σ_γ^2 parameter when it is larger than 0.25.

The data are generated following the conditional Poisson AR(1) model (10.4) for a selected value of σ_γ^2 and longitudinal correlation parameter ρ. In each simulation, three parameters, namely β, σ_γ^2, and ρ are estimated by solving the GQL estimating equations (10.27), (10.36) ($\rho = 0$ based), and moment equation (10.62), respectively. Based on 500 simulations, the simulated mean (SM), simulated standard error (SSE), and the relative bias (RB) defined by

$$\text{RB}(\hat{\beta}) = \frac{[\beta - \text{SM}(\hat{\beta})]}{\text{SE}(\hat{\beta})} \times 100,$$

are reported in Table 10.1.

The results in the table show the GQL approach performs very well in estimating the regression effect β, irrespective of the values of σ_γ^2 and ρ. For example, for the case when $\sigma_\gamma^2 = 0.5$ and $\rho = 0.25$, the GQL approach produces the β estimate as 0.968 for true $\beta = 1.0$ with RB as 42, whereas for larger values of $\sigma_\gamma^2 = 0.75$ and $\rho = 0.75$, the β estimate is found to be slightly worse as 0.955 with RB as 45. This GQL approach also works well for the estimation of the other two parameters, except when both parameters are considerably very large. For example, when σ_γ^2 is large such as $\sigma_\gamma^2 = 0.75$, with a small value for $\rho = 0.25$, the estimates of these parameters appear to be very close to the corresponding true values (0.73 and 0.27, respectively), whereas for the same value of σ_γ^2 with a larger value of $\rho = 0.75$, it produced slightly biased estimates (0.71 and 0.63, respectively) for these parameters, especially for ρ.

10.2.2.2 Simulation Study with $p = 2$ Covariates

Note that whether it is a longitudinal or familial or familial−longitudinal study, the inferences about the model parameters depend on the nature of the covariates. This issue that the covariates play an important role in inferences is clear, for example, from the correlation structure (10.8), where it is seen that the correlations of the responses are functions of the time-dependent covariates. The correlation structure involving the covariates is a key factor in the construction of the estimating equations (10.27) for β, (10.36) for σ_γ^2, and (10.62) for ρ, thus in this section we study the performance of the GQL estimation approach for a case when the model contains more than one time-dependent covariate. As in the last simulation study in Section 10.2.2.1, we consider $K = 100$, $T = 4$, $n_i = 2$, but $p = 2$ as opposed to $p = 1$. Furthermore, we consider the following two time-dependent covariates as in Sutradhar, Jowaheer, and Sneddon (2008). For the first member in a given family, we consider

$$x_{i1t1} = \begin{cases} 1/2 & \text{for } i = 1,\ldots,K/4;\ t = 1,2 \\ 0 & \text{for} i = 1,\ldots,K/4;\ t = 3,4 \\ -1/2 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 1 \\ 0 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 2,3 \\ 1/2 & \text{for } i = K/4+1,\ldots,3K/4;\ t = 4 \\ t/8 & \text{for } i = 3K/4+1,\ldots,K;\ t = 1,\ldots,4, \end{cases},$$

and

$$x_{i1t2} = \begin{cases} (t-2.5)/2 & \text{for } i = 1,\ldots,K/2;\ t = 1,\ldots,4 \\ 0 & \text{for } i = K/2+1,\ldots,K;\ t = 1,2 \\ 1/2 & \text{for } i = K/2+1,\ldots,K;\ t = 3,4. \end{cases}$$

For the second member of the families we consider the first covariate as the binary variable following the distribution

$$Pr(x_{i2t1} = 1) = \begin{cases} 0.3 & \text{for } t = 1 \\ 0.5 & \text{for } t = 2,3 \\ 0.8 & \text{for } t = 4, \end{cases},$$

and the second covariate was chosen as

$$x_{i2t2} = \begin{cases} (t-2.5)/2 & \text{for } i = 1,\ldots,K/2;\ t = 1,\ldots,4 \\ t/2 & \text{for } i = K/2+1,\ldots,K;\ t = 1,\ldots,4. \end{cases}$$

As far as the parameters are concerned, we consider two sets of regression models: M1 with $\beta_1 = \beta_2 = 0.0$ and M2 with $\beta_1 = 1.0$, $\beta_2 = 0.5$. With regard to the values for the familial (σ_γ^2) and longitudinal (ρ) correlation index parameters, we choose them as in the last simulation study.

Now, by using the chosen covariates, we simulate the responses following the conditional autocorrelation model (10.4) for a selected set of parameter values. In each simulation, we then obtain the GQL estimates for the β and σ_γ^2 parameters by using the GQL estimating equations (10.27) and (10.36), respectively. The moment estimate of ρ is obtained by (10.62). The simulations are repeated 500 times. The simulated estimates are reported in Table 10.2.

For both sets of values of β_1 and β_2, the results in Table 10.2 show that the GQL method performs very well in estimating the regression effects. For example, for true $\beta_1 = \beta_2 = 0.0$, the estimates of β_1 and β_2 are found to be 0.006 and -0.010, with corresponding RB 2 and 11 only, when $\sigma_\gamma^2 = 0.75$ and $\rho = 0.75$. This good behavior of the GQL estimates appears to hold for the estimation of the nonzero true values of β_1 and β_2, under various selection for the other two parameter values. Next, for the true value of $\rho = 0.75$, when $\beta_1 = 0$, $\beta_2 = 0$, the correlation estimate was found to be $\hat{\rho} = 0.760$, indicating that the moment estimate of ρ is almost unbiased and hence consistent. This consistency pattern appears to hold for the estimation of this ρ parameter, irrespective of the selected true values for β_1, β_2, and σ_γ^2, as shown in the Table 10.2. As far as the estimation of σ_γ^2 is concerned, the estimates are reasonably unbiased in general except for the cases with a large value of ρ. For example, when $\beta_1 = 1.0$, $\beta_2 = 0.5$, and $\rho = 0.75$, the estimate of σ_γ^2 is found to be 0.464 which is slightly biased for true $\sigma_\gamma^2 = 0.5$, and 0.663, a biased estimate, for true $\sigma_\gamma^2 = 0.75$. This poor performance of the GQL estimation for σ_γ^2 is, however, not surprising, as the 'working' GQL estimating equation (10.36) is constructed based on the weight matrix $\Omega_i(\rho = 0)$ ignoring the longitudinal correlation ρ. Recall that it was demonstrated through a simulation study in Chapter 8 (see Tables

Table 10.2 Simulated mean, simulated standard error, and relative bias of the GQL estimates for the parameters of two nonstationary familial longitudinal regression models, M1: $\beta_1 = 0,\ \beta_2 = 0$; M2: $\beta_1 = 1.0,\ \beta_2 = 0.5$; each with various selected values of the variance component (σ^2_γ), and longitudinal correlation index parameter (ρ); and $K = 100$; $T = 4$; 500 simulations.

				Estimates			
Model	σ^2_γ	ρ	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2_\gamma$	$\hat{\rho}$
M1	0.50	0.25	SM	0.021	−0.004	0.458	0.237
			SSE	0.270	0.080	0.164	0.144
			RB	8	5	26	9
		0.75	SM	−0.023	0.004	0.393	0.734
			SSE	0.193	0.080	0.157	0.077
			RB	12	5	68	21
	0.75	0.25	SM	0.021	−0.005	0.636	0.256
			SSE	0.320	0.099	0.190	0.164
			RB	7	5	60	4
		0.75	SM	0.006	−0.010	0.593	0.760
			SSE	0.260	0.095	0.215	0.063
			RB	2	11	73	16
M2	0.50	0.25	SM	1.013	0.494	0.478	0.259
			SSE	0.242	0.080	0.145	0.158
			RB	5	8	15	6
		0.75	SM	1.006	0.497	0.464	0.746
			SSE	0.221	0.077	0.146	0.083
			RB	3	4	25	5
	0.75	0.25	SM	0.973	0.498	0.670	0.260
			SSE	0.381	0.112	0.162	0.197
			RB	7	2	49	5
		0.75	SM	0.982	0.499	0.663	0.745
			SSE	0.319	0.109	0.171	0.144
			RB	6	1	51	3

8.10 and 8.11) in the context of longitudinal mixed model for count data, that the normality approximation based GQL approach works better than the independence assumption based GQL approach in estimating the variance component of the random effects. This result should hold also in the present familial longitudinal setup. Thus, it is expected that in the present setup, the GQL estimating equation (10.45) constructed using the normality based weight matrix $\Omega_{iN}(\rho)$ will improve the estimate for σ^2_γ. We, however, do not add any new simulations for this, for convenience.

10.2.2.3 Effects of Partial Model Fitting: A Further Simulation Study with $p = 2$ Covariates

Note that the familial longitudinal models considered in this chapter reduce to the simpler longitudinal models discussed in Chapter 6 when $\sigma^2_\gamma = 0$, and they reduce to the simpler familial models in Chapter 4 when longitudinal correlations are zero, that is, when $\rho = 0$, (say). Consequently, when regression parameters are of interest in the familial longitudinal setup, as opposed to fitting the complete familial lon-

Table 10.3 Simulated estimates for a partial longitudinal model when in fact the data came from a familial longitudinal model with regression parameters, M1: $\beta_1 = 0,\ \beta_2 = 0$; M2: $\beta_1 = 1.0,\ \beta_2 = 0.5$; each with various selected values of the variance component (σ^2_γ), and longitudinal correlation index parameter (ρ); and $K = 100$; $T = 4$; 500 simulations.

				Estimates		
Model	σ^2_γ	ρ	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
M1	0.50	0.25	SM	0.094	−0.016	0.528
			SE	0.158	0.051	0.047
			RB	60	31	591
		0.75	SM	−0.013	0.019	0.831
			SE	0.148	0.061	0.032
			RB	9	31	253
	0.75	0.25	SM	0.106	−0.012	0.611
			SE	0.156	0.053	0.046
			RB	68	23	785
		0.75	SM	0.022	−0.014	0.867
			SE	0.154	0.072	0.032
			RB	14	19	366
M2	0.50	0.25	SM	0.984	0.514	0.653
			SE	0.130	0.035	0.050
			RB	12	40	806
		0.75	SM	0.981	0.512	0.913
			SE	0.131	0.044	0.085
			RB	15	27	192
	0.75	0.25	SM	0.964	0.521	0.749
			SE	0.143	0.034	0.044
			RB	25	62	1134
		0.75	SM	0.955	0.521	0.966
			SE	0.129	0.045	0.037
			RB	35	47	584

gitudinal model (10.4, say), one may alternatively attempt to use either one of the simpler models (familial or longitudinal). In order to examine the effect of fitting such simpler but partial models, Sutradhar, Jowaheer, and Sneddon (2008, Sections 4.2 – 4.4) have done further simulation studies on model misspecification effects. We consider here two of their cases and examine the estimation effects of ignoring family effects through using $\sigma^2_\gamma = 0$ and of ignoring longitudinal effects through using $\rho = 0$, when in fact none of them are zero in the familial longitudinal data.

To be specific, in each simulation, the familial longitudinal data are generated as in Section 10.2.2.2 using the conditional AR(1) model (10.4). However, in the first case, we pretend that $\sigma^2_\gamma = 0$, which is equivalent to assuming that there do not exist any families, so that, the data came from all $\sum_{i=1}^{K} n_i = n$ individuals, those who are independent of each other. Thus, in this case, we use the GQL estimating equation (10.27) for β estimation and use (10.62) for ρ estimation. Here both equations are evaluated at $\sigma^2_\gamma = 0$. The simulation results based on 500 simulations are shown in Table 10.3. When the results of Table 10.3 are compared to the corresponding results in Table 10.2, it is clear that both regression and the correlation estimates become bi-

Table 10.4 Simulated estimates for a partial familial model when in fact the data came from a familial longitudinal model with regression parameters, M1: $\beta_1 = 0,\ \beta_2 = 0$; M2: $\beta_1 = 1.0,\ \beta_2 = 0.5$; each with various selected values of the variance component (σ^2_γ), and longitudinal correlation index parameter (ρ); and $K = 100$; $T = 4$; 500 simulations.

Model	Convergent simulations	σ^2_γ	ρ	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2_\gamma$
M1	223	0.50	0.25	SM	0.070	−0.010	0.584
				SE	0.388	0.143	0.159
				RB	18	7	53
	56		0.75	SM	0.236	−0.119	0.675
				SE	0.483	0.192	0.214
				RB	49	62	82
	298	0.75	0.25	SM	0.026	−0.011	0.710
				SE	0.461	0.143	0.161
				RB	6	8	25
	149		0.75	SM	0.083	−0.053	0.757
				SE	0.519	0.195	0.184
				RB	16	27	4
M2	444	0.50	0.25	SM	1.014	0.487	0.495
				SE	0.300	0.100	0.166
				RB	5	13	3
	364		0.75	SM	1.017	0.478	0.545
				SE	0.422	0.141	0.172
				RB	4	16	26
	471	0.75	0.25	SM	0.966	0.493	0.679
				SE	0.396	0.129	0.192
				RB	9	5	37
	422		0.75	SM	0.989	0.500	0.683
				SE	0.460	0.150	0.186
				RB	2	0	36

ased when σ^2_γ is ignored. In particular, the correlation estimates appear to be highly biased with relatively smaller standard errors. Thus, the estimator converges to a wrong value quite often showing the inconsistency of the estimates. For example, consider the case with the smaller value of $\sigma^2_\gamma = 0.50$. When $\beta_1 = 1.0$, $\beta_2 = 0.5$, and $\rho = 0.75$, and σ^2_γ is ignored, the results of Table 10.3 show that the relative biases for the estimates of β_1, β_2, and ρ are 15, 27, and 192, respectively, which are substantially larger than the corresponding relative biases 3, 4, and 5 produced in Table 10.2. These results clearly demonstrate that ignoring random family effects variation has severe consequences mainly on the estimation of ρ which itself is an important parameter in the longitudinal setup.

Next, we pretend that $\rho = 0$, which is equivalent to assuming that the repeated count responses from each and every family member are independent. Thus, in this case, we use the GQL estimating equation (10.27) for β estimation and use (10.36) for ρ estimation. Here both equations are evaluated at $\rho = 0$. The simulation results based on 500 simulations are shown in Table 10.4. This estimation situation appears to be gloomy. This is because as shown in Table 10.4, the iterations did not converge to an estimate for many simulations. For example, when $\beta_1 = 1.0$, $\beta_2 = 0.5$, $\sigma_\gamma^2 = 0.75$, and $\rho = 0.75$, but the estimation is carried out by assuming $\rho = 0.0$, the iterations converged in 422 simulations out of 500 simulations. Note that even though the converged estimates for the regression effects β_1, β_2, and the variance parameter σ_γ^2 appear to be satisfactory as in the complete case reported in Table 10.2, it is, however, not recommended to use $\rho = 0.0$ in estimating other parameters. This is because when $\rho = 0.0$ is used, there will be no guarantee that in practice one can obtain the estimates for other parameters if the data really follow the familial longitudinal model with a nonzero ρ value.

10.3 Analyzing Health Care Utilization Data by Using GLLMM

The complete health care utilization data collected from 36 families of size 4, and 12 families of size 3, for a period of six years from 1985 to 1990, are displayed in Table 6A in the appendix of Chapter 6. The data contain the number of visits to the physician by a member of a family at a given year. Thus, this count (i.e., number of visits), can be denoted appropriately by y_{ijt}, where $i = 1,\ldots,K$, $K = 48$ being the number of independent families; $j = 1,\ldots,n_i$, with n_i as the number of members in the ith family ($n_i = 4$ for $i = 1,\ldots,36$; and $n_i = 3$ for $i = 37,\ldots,48$); and $t = 1,\ldots,T$, $T = 6$ being the number of years. The data also contain information on four covariates, namely (i) gender, (ii) initial number (in 1985) of chronic conditions, (iii) education level, and (iv) age, corresponding to each y_{ijt}. The four-dimensional covariate vector can be represented by $x_{ijt} = [x_{ijt1}, x_{ijt2}, x_{ijt3}, x_{ijt4}]'$. Let $\beta = [\beta_1, \beta_2, \beta_3, \beta_4]'$ denote the effects of the fixed covariate vector x_{ijt} on the response y_{ijt}. Note that there arise the following two types of correlations among the responses in this familial longitudinal setup.

(a) Familial correlations. At a given time t, the responses from any two members, say y_{ijt} and y_{ikt}, under the ith family, will be correlated as they share the common random family effects, say γ_i. This causes familial correlations.

(b) Longitudinal correlations. Conditional on the latent family effect γ_i, the responses collected at two time points, say y_{iju} and y_{ijt}, from the same (jth) member of the ith family will also be correlated. These correlations are referred to as the longitudinal correlations as they take place because of certain dynamic relationships between responses over time.
Note that it is important to take these correlations into account and then compute the regression effects β. Furthermore, in many situations these correlations may also be of primary interest. Now for a complete analysis of this type of data, one may apply a suitable familial longitudinal model introduced in Section 10.1.2. These models, namely conditional AR(1), MA(1), and EQC, belong to a nonstationary autocorrelations class and are simple to implement for their low order. Because the AR(1)

model shows correlation decay over time, which is mostly expected in practice, Sutradhar, Jowaheer, and Sneddon (2008) have fitted this model the familial longitudinal health care utilization data. We discuss it below as an illustration.

Note that a part of this health care dataset in a familial setup was analyzed by Chowdhury and Sutradhar (2009). These authors considered the data from all 48 families for 1985 only. For the covariates: gender (x_{ij1}), the chronic condition (x_{ij2})[CC], education level (x_{ij3})[EL], and age of the individual (x_{ij4}); coded as

$$x_{ij1} = \begin{cases} 0 & \text{female} \\ 1 & \text{male} \end{cases} \qquad x_{ij2} = \begin{cases} 0 & \text{without chronic diseases} \\ 1 & \text{with chronic diseases} \end{cases}$$

$$x_{ij3} = \begin{cases} 0 & \text{less than high school} \\ 1 & \text{high school or above} \end{cases} \qquad x_{ij4} = \text{ exact age of the individual,}$$

their effects β, on the count responses, along with the estimate for familial correlation index parameter (variance of the random effects, σ_γ^2) were reported in Chapter 4 (see Table 4.10 in Section 4.2.8). In a longitudinal setup, Sutradhar (2003) has analyzed a part of the dataset, specifically, the repeated count responses from 144 individuals (members of first 36 families), collected over four time points from 1985 to 1988. By using a reverse code for education level, namely,

$$x_{ij3} = \begin{cases} 1 & \text{less than high school} \\ 0 & \text{high school or above,} \end{cases}$$

but the same codes for the other three covariates as in the familial setup, Sutradhar (2003) has computed the GQL estimates for the components of β, as well as the longitudinal correlation index parameter ρ. These estimates were reported in Chapter 6 (see Table 6.13 in Section 6.7). Note that in this longitudinal setup, it was assumed that 144 individuals were selected independently even though they belong to 36 famililes. Thus, this analysis was done by ignoring the family variation or familial correlation (i.e., by pretending that $\sigma_\gamma^2 = 0$). We remark here that this longitudinal model fitting to the familial longitudinal data by Sutradhar (2003) was done simply for an illustration. Later on Sutradhar, Jowaheer, and Sneddon (2008) have analyzed the same familial longitudinal data by fitting the appropriate familial longitudinal model for count data, namely the conditional AR(1) model $(10.4) - (10.8)$. It is clear from Sutradhar, Jowaheer, and Sneddon (2008) that it is quite important to consider the σ_γ^2 parameter for the health care utilization data. This is because as shown in the following Table 10.5, the sample variance for 144 individuals at a given year for a period of four years appears to be much larger than the corresponding sample mean, indicating over dispersion, that is, the presence of σ_γ^2. To be specific, the sample means and variances appear to reflect the mean and variances shown in (10.8) under the conditional AR(1) model (10.4) (or MA(1) (10.9) or EQC (10.20)); that is,

Table 10.5 Summary statistics of physician visits by 144 members of 36 families at a given year for a period of four years.

	Year			
	1	2	3	4
Average number of visit	3.88	3.75	3.85	4.31
Sample variance	19.65	16.85	18.05	23.24

$$\mu_{ijt} = E[Y_{ijt}] = \exp(x'_{ijt}\beta + \frac{1}{2}\sigma_\gamma^2); \quad \text{var}[Y_{ijt}] = \sigma_{ij,tt} = \mu_{ijt} + [\exp(\sigma_\gamma^2) - 1]\mu_{ijt}^2.$$

Table 10.6 GQL estimates (EST) along with standard errors (SE) (where appropriate) by fitting (a) nonstationary AR(1) model (10.4) – (10.8), and (b) nonstationary longitudinal model (6.44) [re-display from Table 6.13], to the health care utilization data for 36 families each with four members; and (c) familial model (4.1) – (4.2) [re-display from Table 4.10] to all 44 families at one time point, 1985.

GQL Applying to

	Full Model (a)		Partial Model (b)		Partial Model (c)	
Parameters	EST	SE	EST	SE	EST	SE
Gender effect (β_1)	−0.468	0.003	−0.223	0.060	−0.754	0.091
Chronic effect (β_2)	0.331	0.004	0.374	0.072	0.666	0.125
Education effect (β_3)	0.486	0.003	−0.428	0.074	0.434	0.123
Age effect(β_4)	0.024	0.000	0.029	0.001	0.010	0.003
Variance component (σ_γ^2)	1.184	0.314	–	–	0.873	0.409
ρ	0.447	–	0.554	–	–	–

Further note that because the consistent and efficient estimation for β by (10.27), and for σ_γ^2 by (10.36) [or (10.45)], requires the consistent estimate for the longitudinal correlation parameter ρ, we also estimate this later parameter by using the moment estimating equation (10.62). We also compute the standard errors of the estimate of β by (10.29) and of the estimate of σ_γ^2 by (10.44). These estimates (EST) and standard errors (SE) of the estimates are given in Table 10.6. In the same table, for a clear comparison, we also re-display the estimates from Table 6.13 obtained by fitting AR(1) longitudinal model, and the estimates from Table 4.10 obtained by fitting a familial model (to 1985 data).

When the estimates under the full (familial longitudinal) model (a) are compared to those of partial (1985 familial data) model (c), it becomes clear that the family effects have a significant variation that is evident by the estimates for σ_γ^2 under both models (a) and (c). The estimate of σ_γ^2 under full model (a) is computed by using data for four years, whereas it is computed based on 1985 data only under model (c). The estimate under model (a) with smaller standard error as compared to model (c) is naturally more reliable. This large value (1.184) for σ_γ^2 also supports the presence of overdispersion or family variation reflected by the mean and variance comparison in Table 10.5. All estimates under model (a) are improvement over the estimates found under model (3). The estimates of gender and chronic effects appear to be quite different under these two models, whereas education and age effects are almost the same under the two models. The standard errors under model (a) are uniformly smaller than those under model (c). This gain in efficiency can be interpreted as an effect of considering the larger dataset under model (a), or more specifically, as an effect of using the longitudinal correlation in the estimating equation (10.27) for regression effects β and in the estimating equation (10.36) for σ_γ^2. Under the full model (a), the longitudinal correlation estimate was found to be 0.447, a large positive value, indicating that the conditional correlation structure (10.4) – (10.8) plays a significant role in the analysis of the data.

Note that a comparison between the estimates under full model (a) and the longitudinal model (b) show that except for gender effect, the regression estimates are almost the same. This is not surprising as both models are fitted to the same large longitudinal dataset. However, as expected, the standard errors under model (a) are uniformly smaller than the corresponding standard errors under model (b). This efficiency gain can be interpreted as an improvement due to the utilization of the family effects variation under model (a) as compared to the longitudinal model (b) where family variation was treated to be zero, even though it is highly significant.

We now interpret the regression estimates under the familial longitudinal model (a). Because the female was coded as 0 and the male as 1, the large negative value of $\hat{\beta}_1 = -0.468$ indicates that the females pay more visits to the physician as compared to the males. The positive large values of $\hat{\beta}_2 = 0.331$ and $\hat{\beta}_3 = 0.486$ imply that the individuals having a large number of chronic diseases or the individuals with high education pay more visits to the physicians. Similarly, the positive value of $\hat{\beta}_4 = 0.024$ indicates that the subjects in the higher age group appear to pay more visits as compared to the individuals in the lower age group, as expected.

10.4 Some Remarks on Model Identification

In the present familial longitudinal setup, the familial correlations are introduced through the random family effects. As far as the longitudinal correlations are concerned, it has been argued that the repeated data from the family members are most likely to follow one of the conditional (on random family effect) nonstationary low-order autocorrelations such as the AR(1), MA(1), or EQC model, discussed in Section 10.1.2. However, among these three structures, AR(1) perhaps will be the most probable model for repeated data because of the fact that the autocorrelations under this model exhibit a decaying correlation nature as the lag increases. For this reason, for convenience, in Sections 10.2 and 10.3, we have demonstrated how to develop inferences in the conditional AR(1) familial longitudinal model (10.4). More specifically, the covariance structure (10.8) under this AR(1) model was used to develop the GQL estimating equations for β as in (10.27) and for σ_γ^2 as in (10.36) [or (10.45)], and the longitudinal index correlation ρ was estimated by using the moment equation (10.62). Note that in case it is found that the conditional MA(1) or EQC model is more appropriate than AR(1) model, the inferences for these models may easily be developed in the same fashion as that for the AR(1) model. The only difference lies in using the appropriate covariance structure such as (10.13) for the MA(1) model and (10.24) for the EQC model, into the estimating equations (10.27), (10.36) [or (10.45)] and write the appropriate formula for the longitudinal correlation index parameter in a similar way to that of (10.62).

For the identification of the longitudinal correlation structure (assuming that one of the aforementioned three models fit the data) within the familial longitudinal setup, we provide below a few important steps, which are quite similar to the identification steps used in the longitudinal setup in Chapter 6 (see Section 6.5.3).

10.4.1 An Exploratory Identification

Note that because the familial correlations are introduced through the random effects, it remains only to identify the longitudinal correlation structure. Also note that in the familial longitudinal setup, the correlation models are written at the conditional level. That is, conditional on the random effects, the repeated responses of the same individual member in a given family are correlated. To have an approximate idea about the correlation structure, follow the steps below:

Step 1. Use $\gamma_i = 0$ (i.e., $\sigma_\gamma^2 = 0$) for all $i = 1, \ldots, K$, so that $\mu^*_{ijt} = \exp(x'_{ijt}\beta)$.
Step 2. Obtain a 'working' independence assumption based estimate for β by solving (10.27) where now

$$\Sigma_i(\beta, \rho, \sigma_\gamma^2) = A_i = \mathrm{diag}[\mu^*_{i11}, \ldots, \mu^*_{ijt}, \ldots, \mu^*_{in_iT}].$$

Step 3. Compute all lag $\ell = 1, \ldots, T$ correlations by using

$$\hat{\rho}_\ell | \gamma_1 = 0, \ldots, \gamma_K = 0 = \frac{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-\ell} \tilde{y}_{ijt} \tilde{y}_{ij,t+\ell} / (T-\ell) \sum_{i=1}^K n_i}{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T} \tilde{y}^2_{ijt} / T \sum_{i=1}^K n_i}, \tag{10.63}$$

where $\tilde{y}_{ijt} = [y_{ijt} - \mu^*_{ijt}] / \sqrt{\mu^*_{ijt}}$ is evaluated by using β estimate from Step 2.
Step 4. Now, similar to the Gaussian autocorrelation process, if the values of $\hat{\rho}_\ell$ decay as ℓ increases, then decide for the AR(1) covariance structure (10.8); if $\hat{\rho}_1$ is nonzero and for other ℓ, $\hat{\rho}_\ell = 0$, then decide for the MA(1) covariance structure (10.13); if however the values of $\hat{\rho}_\ell$ for all $\ell = 1, \ldots, T-1$, are almost the same, then decide for the EQC covariance structure (10.24).

10.4.2 A Further Improved Identification

Step 1. Compute $\hat{\rho}_\ell$ for all $\ell = 1, \ldots, T-1$, following Section 10.4.1.
Step 2. Use the 'working' estimate of β for β from Section 10.4.1 and compute $\mu^*_{ijt} = \exp(x'_{ijt}\beta)$, and go to Step 3.
Step 3. Evaluate the approximate expectation of $\hat{\rho}_\ell$ under all three possible models as follows.

$$\text{For } AR(1):$$

$$E[\hat{\rho}_\ell | \gamma_1 = 0, \ldots, \gamma_K = 0] \equiv \frac{\rho^\ell}{\sum_{i=1}^K n_i(T-\ell)} \sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-\ell} \left[\frac{\mu^*_{ijt}}{\mu^*_{ij,t+\ell}}\right]^{1/2}$$
$$\text{for } \ell = 1, \ldots, T-1, \tag{10.64}$$

$$\text{For } MA(1):$$

$$E[\hat{\rho}_\ell | \gamma_1 = 0, \ldots, \gamma_K = 0] \equiv \begin{cases} \sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-\ell} \left[\frac{\sum_{v=0}^{t-1}(-\rho)^v \mu^*_{ij,t-v}}{\sqrt{\mu^*_{ijt}\mu^*_{ij,t+\ell}}}\right] & \\ \times \left[\frac{\rho}{\sum_{i=1}^K n_i(T-\ell)}\right] & \text{for } \ell = 1 \\ 0 & \text{otherwise} \end{cases} \tag{10.65}$$

$$\text{For } EQC:$$

$$E[\hat{\rho}_\ell | \gamma_1 = 0, \ldots, \gamma_K = 0] \equiv \frac{\rho}{\sum_{i=1}^K n_i(T-\ell)} \sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-\ell} \left[\frac{\mu^*_{i1}}{\{\mu^*_{it}\mu^*_{i,t+\ell}\}^{\frac{1}{2}}}\right]$$
$$\text{for } \ell = 1, \ldots, T-1, \tag{10.66}$$

Step 4. Compare the pattern of $\hat{\rho}_\ell$ with that of $E[\hat{\rho}_\ell | \gamma_1 = 0, \ldots, \gamma_K = 0]$ from Step 3, for all possible values of ρ, under all three models. Choose the model under which the expected values are in closest agreement with the pattern of $\hat{\rho}_\ell$.

Exercises

10.1. (Section 10.2.1.2.) [First order derivatives of λ_i with respect to σ^2_γ (equation (10.34))]
Notice from (10.31) – (10.33) that for the derivation of $\partial\lambda'_i/\partial\sigma^2_\gamma$ it is sufficient to take the derivatives of

$$\lambda_{ij,tt} = \mu_{ijt} + [\exp(\sigma^2_\gamma)]\mu^2_{ijt}$$

and

$$\lambda_{ij,vt} = \rho^{t-v}\mu_{ijv} + [\exp(\sigma^2_\gamma)]\mu_{ijv}\mu_{ijt}, \ \text{ for } v < t,$$

with respect to σ^2_γ, where $\mu_{ijt} = \exp(x'_{ijt}\beta + \sigma^2_\gamma/2)$. Verify that these derivatives have the formulas:

$$\frac{\partial \lambda_{ij,tt}}{\partial \sigma_\gamma^2} = \frac{\mu_{ijt}}{2} + 2\exp(\sigma_\gamma^2)\mu_{ijt}^2,$$

and

$$\frac{\partial \lambda_{ij,vt}}{\partial \sigma_\gamma^2} = \frac{1}{2}\rho^{t-v}\mu_{ijv} + 2\exp(\sigma_\gamma^2)\mu_{ijt}\mu_{ijv}.$$

10.2. (Section 10.2.1.2.1) [Fourth-order moments when $\rho = 0$ for (10.40)]
For $\mu^*_{ijt} = \exp(x'_{ijt}\beta + \gamma_i)$ as in (10.1) [see also (10.5)], use the conditional Poisson marginal moments

$$E(Y_{ijt}|\gamma_i) = \mu^*_{ijt},\ E(Y_{ijt}^2|\gamma_i) = \mu^*_{ijt} + {\mu^*}^2_{ijt},$$
$$E(Y_{ijt}^3|\gamma_i) = \mu^*_{ijt} + 3{\mu^*}^2_{ijt} + {\mu^*}^3_{ijt},$$

and

$$E(Y_{ijt}^4|\gamma_i) = \mu^*_{ijt} + 7{\mu^*}^2_{ijt} + 6{\mu^*}^3_{ijt} + {\mu^*}^4_{ijt},$$

and verify in a fashion similar to Section 8.2.2 (the longitudinal mixed model setup) that in the present familial longitudinal setup, the fourth-order raw moments at $\rho = 0$ for the jth member of the ith family are given by

$$\begin{aligned}
(a(i))\ E(Y_{ijt}^4) &= \phi_{i,jj,tttt} \\
&= E_{\gamma_i}[E(Y_{ijt}^4|\gamma_i)] \\
&= \mu_{ijt}[1 + 7\mu_{ijt}\exp(\sigma_\gamma^2) + 6\mu_{ijt}^2\exp(3\sigma_\gamma^2) \\
&+ \mu_{ijt}^3\exp(6\sigma_\gamma^2)] \quad (10.67) \\
(a(ii), a(iv))\ E(Y_{iju}^2 Y_{ijt}^2|\rho = 0) &= \phi_{i,jj,uutt} \\
&= E_{\gamma_i}[E(Y_{iju}^2|\gamma_i)E(Y_{ijt}^2|\gamma_i)] \\
&= \mu_{iju}\mu_{ijt}\exp(\sigma_\gamma^2)[1 + \{\mu_{iju} + \mu_{ijt}\}\exp(2\sigma_\gamma^2) \\
&\quad + \mu_{iju}\mu_{ijt}\exp(5\sigma_\gamma^2)] \quad (10.68) \\
(a(iii-1))\ E(Y_{iju}^3 Y_{ijt}|\rho = 0) &= \phi_{i,jj,uuut} \\
&= E_{\gamma_i}[E(Y_{iju}^3|\gamma_i)E(Y_{ijt}|\gamma_i)] \\
&= \mu_{iju}\mu_{ijt}\exp(\sigma_\gamma^2)[1 + 3\mu_{iju}\exp(2\sigma_\gamma^2) \\
&\quad + \mu_{iju}^2\exp(5\sigma_\gamma^2)] \quad (10.69) \\
(a(iii-2))\ E(Y_{iju}^2 Y_{ijv} Y_{ijt}|\rho = 0) &= \phi_{i,jj,uuvt} \\
&= E_{\gamma_i}[E(Y_{iju}^2|\gamma_i)E(Y_{ijv}|\gamma_i)E(Y_{ijt}|\gamma_i)]
\end{aligned}$$

$$
\begin{aligned}
&= \mu_{iju}\mu_{ijv}\mu_{ijt}\exp(3\sigma_\gamma^2) \\
&\quad \times[1+\mu_{iju}\exp(3\sigma_\gamma^2)] \qquad (10.70)\\
(a(v))\ E(Y_{iju}Y_{ijv}Y_{ij\ell}Y_{ijt}|\rho=0) &= \phi_{i,jj,uv\ell t} \\
&= E_{\gamma_i}[E(Y_{iju}|\gamma_i)E(Y_{ijv}|\gamma_i)E(Y_{ij\ell}|\gamma_i)E(Y_{ijt}|\gamma_i)] \\
&= \mu_{iju}\mu_{ijv}\mu_{ij\ell}\mu_{ijt}\exp(6\sigma_\gamma^2). \qquad (10.71)
\end{aligned}
$$

10.3. (Section 10.2.1.2.1) [Fourth-order moments when $\rho = 0$ for (10.41)]
Verify in a fashion similar to Exercise 10.2 that the fourth-order raw moments at $\rho = 0$ for the jth and kth members of the ith family are given by

$$
\begin{aligned}
(b(i)\ E(Y_{iju}^2 Y_{ikt}^2|\rho=0) &= \phi_{i,jk,uutt} \\
&= E_{\gamma_i}[E(Y_{iju}^2|\gamma_i)E(Y_{ikt}^2|\gamma_i)] \\
&= \mu_{iju}\mu_{ikt}\exp(\sigma_\gamma^2)[1+\{\mu_{iju}+\mu_{ikt}\}\exp(2\sigma_\gamma^2) \\
&\quad +\mu_{iju}\mu_{ikt}\exp(5\sigma_\gamma^2)] \qquad (10.72)\\
(b(ii)\ E(Y_{iju}^2 Y_{ikv}Y_{ikt}|\rho=0) &= \phi_{i,jk,uuvt} \\
&= E_{\gamma_i}[E(Y_{iju}^2|\gamma_i)E(Y_{ikv}|\gamma_i)E(Y_{ikt}|\gamma_i)] \\
&= \mu_{iju}\mu_{ikv}\mu_{ikt}\exp(3\sigma_\gamma^2) \\
&\quad \times[1+\mu_{iju}\exp(3\sigma_\gamma^2)] \qquad (10.73)\\
(b(iii))\ E(Y_{iju}Y_{ijv}Y_{ik\ell}Y_{ikt}|\rho=0) &= \phi_{i,jk,uv\ell t} \\
&= E_{\gamma_i}[E(Y_{iju}|\gamma_i)E(Y_{ijv}|\gamma_i)E(Y_{ik\ell}|\gamma_i)E(Y_{ikt}|\gamma_i)] \\
&= \mu_{iju}\mu_{ijv}\mu_{ik\ell}\mu_{ikt}\exp(6\sigma_\gamma^2). \qquad (10.74)
\end{aligned}
$$

References

1. Breslow, N, E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. . *J. Amer. Stat. Assoc.,* 88, 9 – 25.
2. Breslow, N. E. & Lin, X. (1995). Bias correction in generalized linear models with a single component of dispersion. *Biometrika,* 82, 81 – 92.
3. Chowdhury, M. R. I. & Sutradhar, B. C. (2009). Generalized quasilikelihood versus hierarchial likelihood inferences in generalized linear mixed models for count data. *Sankhya B,* 71, 1 – 15.
4. Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.,* 93, 720 – 729.
5. Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.
6. Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statist. Sci.,* 18, 377 – 393.
7. Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhya B,* 66, 261 – 289.

8. Sutradhar, B. C. (2010). Inferences in generalized linear longitudinal mixed models. *Canad. J. Statist.,* 38, 174 – 196.
9. Sutradhar, B. C. & Jowaheer, V. (2003). On familial longitudinal Poisson mixed models with gamma random effects. *J. Multivariate Anal.,* 87, 398 – 412.
10. Sutradhar, B. C., Jowaheer, V., & Sneddon, G. (2008). On a unified generalized quasi-likelihood approach for familial-longitudinal non-stationary count data. *Scand. J. Statist.,* 35, 597 – 612.

Chapter 11
Familial Longitudinal Models for Binary Data

In the familial longitudinal setup, binary responses along with a set of multidimensional time-dependent covariates are collected from the members of a large number of independent families. For example, in a clinical study, the asthma status of each of the family members of a large number of independent families may be collected every year over a period of four years. Also, the covariates such as gender, age, education level, and life style of the individual member may be collected. In this setup, it is likely that the responses from the members of the same family at a given year will be correlated. This is due to the fact that every member of the family shares certain common family effects which are latent or invisible. Also, the repeated asthma status collected over several years will be longitudinally correlated. It is of interest to take these two types of familial and longitudinal correlations into account and then find the effects of the covariates on the responses. Two types of familial longitudinal models, namely conditional linear and nonlinear models are introduced in Sections 11.1 and 11.3, to analyze this type of two-way correlated binary data. Note that these models for familial longitudinal binary data would be a generalization of either familial models discussed in Chapter 5 or longitudinal models discussed in Chapter 7. In the conditional linear model setup, it is assumed that conditional on random family effects, the repeated binary responses from a member of a given family follow a LDCP (linear dynamic conditional probability) model as in Section 7.4. We refer to such a model as the linear dynamic conditional-conditional probability (LDCCP) model. This model along with the estimation of the parameters is discussed in Section 11.1. An illustration of the estimation methodology is given in Section 11.2 to analyze the Waterloo Smoking Prevention Project-3 (WSPP3) data.

In the conditional nonlinear model setup, it is assumed that conditional on the random family effects, the repeated binary responses from a member of a given family follow a BDL (binary dynamic logit) model as in Section 7.7.2. In the longitudinal setup, this random effects based BDL model was referred to as the binary dynamic mixed logit (BDML) (see Section 9.2) model. In the familial longitudinal setup, we refer to this family based BDML model as the FBDML model. This model along with the estimation of the regression and the dynamic dependence parameters is discussed in Section 11.3. Both LDCCP and FBDML models are discussed in

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_11,

Sections 11.1 and 11.3, respectively, and involve one variance component due to a single random family effect.

Let y_{ijt} denote the binary response for the jth $(j = 1,\ldots,n_i)$ individual on the ith $(i = 1,\ldots,K)$ family/cluster at a given time t $(t = 1,\ldots,T)$. Also, let $x_{ijt} = (x_{ijt1},\ldots,x_{ijtp})'$ denote the p covariates associated with the response y_{ijt}, and β denote the effect of the covariate vector x_{ijt} on y_{ijt}. Similar to the familial longitudinal models for count data introduced in the last chapter, the binary longitudinal responses of any two family members will also exhibit both familial and longitudinal correlations. As an extension of the binary longitudinal models discussed in Chapter 7, one may write linear or nonlinear conditional probability models as in Sections 11.1 and 11.3, to accommodate both familial and longitudinal correlations. Note that these binary models are quite different from the familial longitudinal models given in Chapter 10 for count data.

11.1 LDCCP Models

11.1.1 Conditional-Conditional (CC) AR(1) Model

Suppose that conditional on the random family effect γ_i, the repeated binary responses from the jth $(j = 1,\ldots,n_i)$ member of the ith family follow the LDCP (linear dynamic conditional probability) model of AR(1) form given in (7.70). That is, the CC AR(1) model has the form:

$$\begin{aligned} Pr[Y_{ij1} = 1|\gamma_i] &= \pi^*{}_{ij1} \\ Pr[Y_{ijt} = 1|y_{ij,t-1},\gamma_i] &= \pi^*{}_{ijt} + \rho(y_{ij,t-1} - \pi^*{}_{i,t-1}), \end{aligned} \tag{11.1}$$

for $j = 1,\ldots,n_i$; $t = 2,\ldots,T$. In (11.1), $\pi^*{}_{ijt} = \exp(x'_{ijt}\beta + \gamma_i)/[1+\exp(x'_{ijt}\beta + \gamma_i)]$ for all $j = 1,\ldots,n_i$, $t = 1,\ldots,T$.

11.1.1.1 Conditional Mean, Variance, and Correlation Structure

Conditional on the random family effects γ_i, the linear dynamic probability model (11.1) yields the conditional means and the variances, for the jth member of the ith family at a time point t, as

$$\begin{aligned} E(Y_{ijt}|\gamma_i) &= \pi^*_{ijt} \\ \text{var}(Y_{ijt}|\gamma_i) &= \sigma^*_{i,jj,tt} = \pi^*_{ijt}(1-\pi^*_{ijt}), \end{aligned} \tag{11.2}$$

for $t = 1,\ldots,T$. Next, for $u < t$, by using the model relationship (11.1), one may compute the conditional covariance between y_{iju} and y_{ijt} as

$$\text{cov}[(Y_{iju},Y_{ijt})|\gamma_i] = \rho^{t-u}\sigma^*_{i,jj,uu}, \tag{11.3}$$

yielding the conditional correlations as

$$corr[(Y_{iju},Y_{ijt})|\gamma_i] = \begin{cases} \rho^{t-u}\left[\frac{\sigma^*_{i,jj,uu}}{\sigma^*_{i,jj,tt}}\right]^{1/2}, & \text{for } u < t \\ \rho^{u-t}\left[\frac{\sigma^*_{i,jj,tt}}{\sigma^*_{i,jj,uu}}\right]^{1/2}, & \text{for } u > t \end{cases}. \tag{11.4}$$

As far as the longitudinal correlations between two members of a family are concerned, we assume that at any two time points, the responses of any two members are conditionally independent. In notation,

$$\text{cov}[\{Y_{iju},Y_{ikt}\}|\gamma_i] = 0, \text{ for } j \neq k. \tag{11.5}$$

11.1.1.2 Unconditional Mean, Variance, and Correlation Structure

The unconditional means, variances, and the covariances may be computed by using the following formulas

$$\begin{aligned} E[Y_{ijt}] &= E[Y^2_{ijt}] = E_{\gamma_i}E[Y_{ijt}|\gamma_i] \\ \text{var}[Y_{ijt}] &= E_{\gamma_i}[\text{var}\{Y_{ijt}|\gamma_i\}] + \text{var}_{\gamma_i}[E\{Y_{ijt}|\gamma_i\}] \\ \text{cov}[Y_{iju},Y_{ikt}] &= E_{\gamma_i}\text{cov}[\{Y_{iju},Y_{ikt}\}|\gamma_i] + \text{cov}_{\gamma_i}[E(Y_{iju}|\gamma_i),E(Y_{ikt}|\gamma_i)], \end{aligned} \tag{11.6}$$

where the conditional means, variances and covariances are given by the equations from (11.2) to (11.5). Based on the assumption that $\gamma_i \overset{iid}{\sim} N(0,\sigma^2_\gamma)$, and by using $\gamma^*_i = \gamma_i/\sigma_\gamma$ with $g_N(\gamma^*_i|1)$ as the standard normal density, the unconditional first moments given in (11.6) may be simplified as

$$\begin{aligned} E[Y_{ijt}] = \pi_{ijt}(\beta,\sigma^2_\gamma) &= \int \pi^*_{ijt}(\gamma^*_i)g_N(\gamma^*_i|1)d\gamma^*_i \\ &= \sum_{v_i=0}^{V}[\pi^*_{ijt}(v_i)]\binom{V}{v_i}(1/2)^{v_i}(1/2)^{V-v_i}, \\ &= \pi^{(b)}_{ijt}(\beta,\sigma^2_\gamma), \text{ (say) [binomial approximation]}, \end{aligned} \tag{11.7}$$

where for a known reasonably big V such as $V = 5$, $v_i \sim \text{binomial}(V,1/2)$, and hence it has relation to γ^*_i as

$$\gamma^*_i = \frac{v_i - V(1/2)}{V(1/2)(1/2)}$$

[see (9.10)], so that

$$\pi^*_{ijt}(v_i) = \pi^*_{ijt}(\gamma^*_i)|_{\gamma^*_i=(v_i-V(1/2))/[V(1/2)(1/2)]}.$$

It also follows that the unconditional variance has the formula

$$\text{var}[Y_{ijt}] = \sigma_{i,jj,tt}(\beta, \sigma_\gamma^2) = \pi_{ijt}(\beta, \sigma_\gamma^2)(1 - \pi_{ijt}(\beta, \sigma_\gamma^2)). \tag{11.8}$$

Next, for $j = k$, the unconditional covariances in (11.6) may be simplified as

$$\begin{aligned}\text{cov}[Y_{iju}, Y_{ijt}] = \sigma_{i,jj,ut}(\beta, \sigma_\gamma^2, \rho) &= \rho^{t-u}\left[\pi_{iju} - \int {\pi^*}^2_{iju}(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^*\right] \\ &\quad + \left[\int \pi^*_{iju}(\gamma_i^*)\pi^*_{ijt}(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* - \pi_{iju}\pi_{ijt}\right] \\ &= \rho^{t-u}\left[\pi_{iju} - \pi_{ijj,uu}\right] \\ &\quad + \left[\pi_{ijj,ut} - \pi_{iju}\pi_{ijt}\right], \end{aligned} \tag{11.9}$$

(say). Thus, in general, for all j, k, the unconditional covariances have the formulas

$$\text{cov}(Y_{iju}, Y_{ikt}) = \begin{cases} \rho^{t-u}\left[\pi_{iju} - \pi_{ijj,uu}\right] + \left[\pi_{ijj,ut} - \pi_{iju}\pi_{ijt}\right] & \text{for } k = j; u < t \\ \left[\pi_{ijk,ut} - \pi_{iju}\pi_{ikt}\right] & \text{for } k \neq j; u \leq t, \end{cases} \tag{11.10}$$

where

$$\begin{aligned}\pi_{ijk,ut} &= \int \pi^*_{iju}(\gamma_i^*)\pi^*_{ikt}(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* \\ &= \pi^{(b)}_{ij,ut},\end{aligned}$$

yielding the unconditional AR(1) correlation structure as

$$\text{corr}[Y_{iju}, Y_{ikt}] = \begin{cases} \frac{\rho^{t-u}\left[\pi_{iju} - \pi_{ijj,uu}\right] + \left[\pi_{ijj,ut} - \pi_{iju}\pi_{ijt}\right]}{[\pi_{iju}(\beta,\sigma_\gamma^2)(1-\pi_{iju}(\beta,\sigma_\gamma^2))\pi_{ijt}(\beta,\sigma_\gamma^2)(1-\pi_{ijt}(\beta,\sigma_\gamma^2))]^{1/2}} & \text{for } k = j; u < t \\ \frac{\left[\pi_{ijk,ut} - \pi_{iju}\pi_{ikt}\right]}{[\pi_{iju}(\beta,\sigma_\gamma^2)(1-\pi_{iju}(\beta,\sigma_\gamma^2))\pi_{ikt}(\beta,\sigma_\gamma^2)(1-\pi_{ikt}(\beta,\sigma_\gamma^2))]^{1/2}} & \text{for } k \neq j; u \leq t. \end{cases} \tag{11.11}$$

11.1.2 CC MA(1) Model

Conditional on the random family effects γ_i, we can follow the MA(1) model given in Section 7.4.2, to construct the desired MA(1) model in the familial longitudinal set up. To be specific, we consider

$$\begin{aligned} Pr[Y_{ij1} = 1|\gamma_i] &= {\pi^*}_{ij1} \\ Pr[Y_{ijt} = 1|d_{ijt}, d_{ij,t-1}, \gamma_i] &= d_{ijt}|\gamma_i + \rho d_{ij,t-1}|\gamma_i, \end{aligned} \tag{11.12}$$

for $j = 1, \ldots, n_i$; $t = 2, \ldots, T$. In (11.12), the d_{ijt}s are independently distributed with mean ξ^*_{ijt} and variance η_{ijt}; that is,

$$d_{ijt} \stackrel{id}{\sim} \left[\xi^*_{ijt} = \sum_{v=0}^{t-1} (-\rho)^v \pi^*_{ij,t-v},\ \eta_{it} = [\frac{\sum_{v=0}^{t-1}(-\rho)^v \pi^*_{ij,t-v}}{\sum_{v=0}^{t-1}(-\rho)^v}][1 - \frac{\sum_{v=0}^{t-1}(-\rho)^v \pi^*_{ij,t-v}}{\sum_{v=0}^{t-1}(-\rho)^v}] \right],$$

for all $t = 1, \ldots, T$, and where

$$\pi^*{}_{ijt} = \exp(x'_{ijt}\beta + \gamma_i)/[1 + \exp(x'_{ijt}\beta + \gamma_i)],$$

for all $j = 1, \ldots, n_i,\ t = 1, \ldots, T$. It then follows that the conditional means, variances, and covariances are given by

$$\begin{aligned} E[Y_{ijt}|\gamma_i] &= \pi^*_{ijt} \\ \text{var}[Y_{ijt}|\gamma_i] &= \pi^*_{ijt}[1 - \pi^*_{ijt}] \\ \text{cov}[(Y_{iju}, Y_{ijt})|\gamma_i] &= \begin{cases} \rho \left[[\frac{\sum_{v=0}^{u-1}(-\rho)^v \pi^*_{ij,u-v}}{\sum_{v=0}^{u-1}(-\rho)^v}][1 - \frac{\sum_{v=0}^{u-1}(-\rho)^v \pi^*_{i,u-v}}{\sum_{v=0}^{u-1}(-\rho)^v}] \right] & \text{for } t - u = 1 \\ 0 & \text{for } |t - u| > 1 \end{cases} \end{aligned} \tag{11.13}$$

Furthermore, similar to the AR(1) case (11.5), for any two members $j \neq k$, under the ith family, we assume that

$$\text{cov}[\{Y_{iju}, Y_{ikt}\}|\gamma_i] = 0, \ \text{for } j \neq k.$$

Note that by using the formulas in (11.6), one may then compute the unconditional means, variances, and covariances, under this MA(1) model.

11.1.3 CC EQC Model

To construct the EQC model in the familial longitudinal setup, we follow the EQC model (7.80), but conditional on the random family effect. Thus, we write

$$Pr[Y_{ijt} = 1|y_{ij0}, \gamma_i] = \pi^*{}_{ijt} + \rho(y_{ij0} - \pi^*{}_{ij1}), \ \text{for } j = 1, \ldots, n_i;\ t = 1, \ldots, T, \tag{11.14}$$

where y_{ij0} is an initial unobservable binary response for the jth member of the ith family, with its mean π^*_{ij1}, which is also the mean of y_{ij1}.

Now by using (11.14), we can write the conditional means, variances, and covariances as

$$\begin{aligned} E[Y_{ijt}|\gamma_i] &= \pi^*_{ijt} \\ \text{var}[Y_{ijt}|\gamma_i] &= \pi^*_{ijt}[1 - \pi^*_{ijt}] \\ \text{cov}[(Y_{iju}, Y_{ijt})|\gamma_i] &= \rho^2 \pi^*_{ij1}[1 - \pi^*_{ij1}]. \end{aligned} \tag{11.15}$$

For any two members, we make the same conditional assumption, that is,

$$\mathrm{cov}[\{Y_{iju}, Y_{ikt}\}|\gamma_i] = 0, \text{ for } j \neq k,$$

similar to that of the AR(1) (11.5) and MA(1) models.

One may then derive the formulas for the unconditional means, variances, and covariances by applying (11.15) to (11.6).

11.1.4 Estimation of the AR(1) Model Parameters

Note that among the three LDCCP models: AR(1), MA(1), and EQC; the AR(1) model exhibits decay in longitudinal lag correlations, which is expected in practice for most of the familial longitudinal data. We now, for convenience, provide the estimation of the parameters of the AR(1) LDCCP model. The parameters of the other two models may be estimated similarly. If a model identification issue arises, this can be done in a similar way to that of the familial longitudinal count data case discussed in the last chapter (see Section (10.4.1)).

Further note that once we complete a brief discussion in this section, on the estimation of parameters for the AR(1) LDCCP model, we illustrate the estimation methodology in Section 11.2, with real life data discussed by Sutradhar and Farrell (2004), under a special familial longitudinal model with stationary autocorrelation structure (appropriate for time-independent covariates) that accommodates all three longitudinal correlation models.

We now turn back to the estimation of the parameters of the AR(1) LDCCP model.

11.1.4.1 GQL Estimation of Regression Parameter β

The GQL estimating equation for β has the same form as that under the familial longitudinal count data model. By representing the mean vector of the binary responses under the ith family with π_i, following (10.27), this equation may be written as

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1} (y_i - \pi_i) = 0, \tag{11.16}$$

where

$$y_i = (y_{i1}', \ldots, y_{ij}', \ldots, y_{in_i}')'$$

denote the $n_i T \times 1$ vector of binary responses for the ith family, with

$$y_{ij} = (y_{ij1}, \ldots, y_{ijt}, \ldots, y_{ijT})',$$

y_{ijt} being the binary response recorded at time t $(t = 1, \ldots, T)$, from the jth $(j = 1, \ldots, n_i)$ member of the ith $(i = 1, \ldots, K)$ family. The $n_i T \times 1$ unconditional mean vector of π_i may be expressed as

$$\pi_i(\beta,\sigma_\gamma^2) = (\pi'_{i1}(\beta,\sigma_\gamma^2),\ldots,\pi'_{ij}(\beta,\sigma_\gamma^2),\ldots,\pi'_{in_i}(\beta,\sigma_\gamma^2))', \tag{11.17}$$

where

$$\pi_{ij}(\beta,\sigma_\gamma^2) = (\pi_{ij1}(\beta,\sigma_\gamma^2),\ldots,\pi_{ijt}(\beta,\sigma_\gamma^2),\ldots,\pi_{ijT}(\beta,\sigma_\gamma^2))'$$

is the $T\times 1$ vector with π_{ijt} as its general element. The formula for this general element is given by (11.7).

Also, in (11.16), Σ_i is the $n_iT\times n_iT$ unconditional variance−covariance matrix of y_i, which has the same form as given by (10.26) under the count data model, but, the elements of the block diagonal matrix $\Sigma_{ijj}(\beta,\sigma_\gamma^2,\rho)$ for all $j=1,\ldots,n_i$, are now computed by using the formulas for the unconditional variances and longitudinal covariances for the jth member, given by (11.8)-(11.10), whereas the elements for the off-diagonal matrices $\Sigma_{ijk}(\beta,\sigma_\gamma^2)$ for all $j\neq k$, $j,k=1,\ldots,n_i$, are computed by (11.10).

Furthermore, because $\pi_{ij}=(\pi_{ij1},\ldots,\pi_{ijt},\ldots,\pi_{ijT})'$ with π_{ijt} defined as in (11.7), the derivative of π_i with respect to β' requires the differentiation of π_{ijt} with respect to β. To be specific, $\partial\pi_{ijt}/\partial\beta$ is the $p\times 1$ vector given as

$$\begin{aligned}\frac{\partial\pi_{ijt}}{\partial\beta} &= \int\frac{\partial\pi^*_{ijt}(\gamma_i^*)}{\partial\beta}g_N(\gamma_i^*|1)d\gamma_i^*\\ &= x_{ijt}\sum_{v_i=0}^{V}[\pi^*_{ijt}(\gamma_i^*)\{1-\pi^*_{ijt}(\gamma_i^*)\}]\binom{V}{v_i}(1/2)^{v_i}(1/2)^{V-v_i},\\ &= x_{ijt}[\pi^{(b)}_{ijt}(\beta,\sigma_\gamma^2)-\pi^{(b)}_{ijj,tt}(\beta,\sigma_\gamma^2)],\end{aligned} \tag{11.18}$$

where x_{ijt} is the $p\times 1$ vector of all covariates for the jth individual under the ith family at time t. In (11.18), for

$$\pi^*_{ijt}(v_i) = \pi^*_{ijt}(\gamma_i^*)|_{\gamma_i^*=(v_i-V(1/2))/[V(1/2)(1/2)]}$$

and also

$$\pi_{i,jj,tt}(\beta,\sigma_\gamma^2) = E_{\gamma_i^*}[\pi^{*2}_{ijt}]$$

[see (9.6) for a similar notation].

Let $\hat\beta_{GQL}$ denote the GQL estimator of β, obtained by solving the estimating equation (11.16). This estimator is consistent, and it is highly efficient as the GQL estimating equation is unbiased as well as the weight matrix Σ_i is the true covariance matrix of y_i. Furthermore, by using the multivariate central limit theorem [see Mardia, Kent and Bibby (1979, p.51)], similar to (10.29) for count data, one may show that $K^{1/2}(\hat\beta_{GQL}-\beta)$ has an asymptotic normal distribution, as $K\to\infty$, with mean zero and with covariance matrix given by

$$K\left(\sum_{i=1}^{K}\frac{\partial\pi_i'}{\partial\beta}\Sigma_i^{-1}\frac{\partial\pi_i}{\partial\beta'}\right)^{-1}, \tag{11.19}$$

where Σ_i is the aforementioned covariance matrix of the binary response vector y_i for the ith family.

Remark that the computation of the estimate of β by (11.16) requires the estimates for σ_γ^2 and ρ. These parameters are estimated in the following two sections.

11.1.4.2 GQL Estimation of Familial Correlation Index Parameter σ_γ^2

Similar to the count data case (see Section 10.2.1.2), consider

$$u_{ij} = [u'_{ij(s)}, u'_{ij(p)}]' \tag{11.20}$$

as the $T(T+1)/2$-dimensional combined vector of squares and pairwise products for the jth $(j = 1, \ldots, n_i)$ member of the ith $(i = 1, \ldots, K)$ family, where

$$\begin{aligned} u_{ij(s)} &= [y_{ij1}^2, \ldots, y_{ijt}^2, \ldots, y_{ijT}^2]' : T \times 1 \\ u_{ij(p)} &= [y_{ij1}y_{ij2}, \ldots, y_{ijt}y_{ijv}, \ldots, y_{ij(T-1)}y_{ijT}]' : \frac{T(T-1)}{2} \times 1. \end{aligned}$$

Note, however, that unlike in the count data case, here for the binary responses

$$\begin{aligned} u_{ij(s)} &= [y_{ij1}^2, \ldots, y_{ijt}^2, \ldots, y_{ijT}^2]' : T \times 1 \\ &= [y_{ij1}, \ldots, y_{ijt}, \ldots, y_{ijT}]' \\ &= y_{ij}. \end{aligned} \tag{11.21}$$

Next, we write the $n_iT(T+1)/2$-dimensional vector of squares and distinct products for all n_i individuals in the ith family. Let u_i denote this vector and λ_i be its mean. That is,

$$\begin{aligned} u_i &= [u'_{i1}, \ldots, u'_{ij}, \ldots, u'_{in_i}]' \\ \lambda_i &= [\lambda'_{i1}, \ldots, \lambda'_{ij}, \ldots, \lambda'_{in_i}]', \end{aligned} \tag{11.22}$$

where

$$\lambda_{ij} = [\lambda'_{ij(s)}, \lambda'_{ij(p)}]',$$

with

$$\begin{aligned} \lambda_{ij(s)} &= [E(Y_{ij1}^2), \ldots, E(Y_{ijt}^2), \ldots, E(Y_{ijT}^2)]' \\ &= [\lambda_{ij,11}, \ldots, \lambda_{ij,tt}, \ldots, \lambda_{ij,TT}]' && (11.23) \\ \lambda_{ij(p)} &= [E(Y_{ij1}Y_{ij2}), \ldots, E(Y_{ijv}Y_{ijt}), \ldots, E(Y_{ij(T-1)}Y_{ijT})]', \\ &= [\lambda_{ij,12}, \ldots, \lambda_{ij,vt}, \ldots, \lambda_{ij,T-1,T}]', && (11.24) \end{aligned}$$

where

$$\lambda_{ij,tt} = \pi_{ijt}$$

and for $v < t$,

$$\lambda_{ij,vt} = \rho^{t-v}[\pi_{ijv} - \pi_{ijj,vv}] + \pi_{ijj,vt},$$

by (11.6) and (11.9), with

$$\begin{aligned}
\pi_{ijt} &= \int \pi^*_{ijt}(\gamma^*_i) g_N(\gamma^*_i|1) d\gamma^*_i = \pi^{(b)}_{ijt} \\
\pi_{ijj,tt} &= \int \pi^{*2}_{ijt}(\gamma^*_i) g_N(\gamma^*_i|1) d\gamma^*_i = \pi^{(b)}_{ijj,tt} \\
\pi_{ijj,vt} &= \int \pi^*_{ijv}(\gamma^*_i)\pi^*_{ijt}(\gamma^*_i) g_N(\gamma^*_i|1) d\gamma^*_i = \pi^{(b)}_{ijj,vt},
\end{aligned} \tag{11.25}$$

using the binomial approximation notation from (11.7).

Let $\Omega_i = \text{cov}(U_i)$, and $(\partial\lambda'_i(\beta,\sigma^2_\gamma,\rho))/\partial\sigma^2_\gamma$ be the first derivative vector of λ_i (11.22) with respect to σ^2_γ. Now by computing Ω_i and these derivatives appropriate under the present binary familial longitudinal model, one may obtain, similar to the count data case, the GQL estimate of σ^2_γ either by solving

$$\sum_{i=1}^{K} \frac{\partial\lambda'_i(\beta,\sigma^2_\gamma,\rho)}{\partial\sigma^2_\gamma} \Omega_i^{-1}(\beta,\sigma^2_\gamma,\rho=0)[u_i - \lambda_i(\beta,\sigma^2_\gamma,\rho)] = 0, \tag{11.26}$$

[see (10.36)] or normal approximation based estimating equation

$$\sum_{i=1}^{K} \frac{\partial\lambda'_i(\beta,\sigma^2_\gamma,\rho)}{\partial\sigma^2_\gamma} \Omega_{iN}^{-1}(\beta,\sigma^2_\gamma,\rho)[u_i - \lambda_i(\beta,\sigma^2_\gamma,\rho)] = 0, \tag{11.27}$$

[see (10.45)]. The formulas for the elements of $(\partial\lambda'_i(\beta,\sigma^2_\gamma,\rho))/\partial\sigma^2_\gamma$ are available in Exercise 11.1. The construction of $\Omega_i(\beta,\sigma^2_\gamma,\rho=0)$ and $\Omega_{iN}(\beta,\sigma^2_\gamma,\rho)$ matrices, is given below.

Construction of $\Omega_i(\beta,\sigma^2_\gamma,\rho=0) \equiv \Omega_i(I)$

Recall that the structure of Ω_i is given by (10.35) under the familial longitudinal count data model. This structure remains the same for the binary data. Now, to construct the $\Omega_i(I)$ matrix as a substitute of the Ω_i matrix defined in (10.35), it is sufficient to compute $\Omega_{ijj}(I)$ for all $j = 1,\ldots,n_i$, and $\Omega_{ijk}(I)$ for all $j \neq k, j,k = 1,\ldots,n_i$. Note that when it is pretended that $\rho = 0$, it follows from (11.5) that for $v \neq t$,

$$\text{corr}\{(y_{ijv}, y_{ikt})|\gamma_i\} = 0, \text{ for all } j = k;\ j \neq k. \tag{11.28}$$

Now for the computation of the elements of $\Omega_{ijj}(I)$ and $\Omega_{ijk}(I)$ matrices, we use the conditional independence assumption (11.28) and derive the elements of these two matrices.

Note that for the computation of the elements of the $\Omega_{ijj}(I)$, it is sufficient to compute the formulas for

$$\begin{aligned}
(a):\ & (i)\ \text{var}[Y_{ijt}],\ (ii)\ \text{cov}[Y_{ijt}, Y_{ij\ell}],\ (iii)\ \text{cov}[Y_{ijt}, Y_{ij\ell}Y_{iju}],\ (iv)\ \text{var}[Y_{ijt}Y_{iju}], \text{ and} \\
& (v)\ \text{cov}[Y_{ijt}Y_{iju}, Y_{ij\ell}Y_{ijv}].
\end{aligned} \tag{11.29}$$

Similarly, for the computation of the elements of the $\Omega_{ijk}(I)$, it is sufficient to compute the formulas for

$$(b): \ (i)\ \mathrm{cov}[Y_{iju}, Y_{ikt}],\ (ii)\ \mathrm{cov}[Y_{ijt}, Y_{ik\ell}Y_{iku}],\ (iii)\ \mathrm{cov}[Y_{ijt}Y_{iju}, Y_{ik\ell}Y_{ikv}]. \quad (11.30)$$

These variance and covariances under (a) and (b) may be computed from the formulas available in Exercises 11.2 and 11.3, respectively.

Construction of $\Omega_{iN}(\beta, \sigma_\gamma^2, \rho)$

To construct this normality assumption based fourth-order moment matrix, it is sufficient to construct the normality (N) assumption based two general matrices, $\Omega_{ijj}(N)$ and $\Omega_{ijk}(N)$. Note that to construct the $\Omega_{iN}(\cdot)$ matrix under the normality assumption, one pretends that the

$$y_i = (y'_{i1}, \ldots, y'_{ij}, \ldots, y'_{in_i})' \ : n_iT \times 1$$

binary response vector follows the n_iT-dimensional multivariate normal vector but with true binary mean vector π_i (11.17) and binary AR(1) correlation structure based covariance matrix

$$\Sigma_i(\beta, \sigma_\gamma^2, \rho) = (\sigma_{i,jk,ut}) : \ n_iT \times n_iT, \quad (11.31)$$

where the formulas for

$$\sigma_{i,jk,ut} = \mathrm{cov}[Y_{iju}, Y_{ikt}]$$

for all $j, k = 1, \ldots, n_i$, and $u, t = 1, \ldots, T$, are given by (11.10).

Construction of $\Omega_{ijj}(N)$

Recall that Ω_{ijj} is the $\{T(T+1)/2\} \times \{T(T+1)/2\}$ covariance matrix of

$$u_{ij} = (u'_{ij(s)}, u'_{ij(p)})',$$

$u_{ij(s)}$ being the $T \times 1$ vector of squares of the elements of $y_{ij} = (y_{ij1}, \ldots, y_{ijT})'$, and $u_{ij(p)}$ is the $\{T(T-1)/2 \times 1\}$ vector of distinct pairwise products of the elements of y_{ij}. Now by using the marginal property of the multivariate normal distribution, we write by following (11.31) that

$$y_{ij} = [y_{ij1}, \ldots, y_{ijt}, \ldots, y_{ijT}] \sim N_T(\mu_{ij}, \Sigma_{ijj}), \quad (11.32)$$

where

$$\Sigma_{ijj} = (\sigma_{i,jj,ut})$$

with $\sigma_{i,jj,ut} = \mathrm{cov}[Y_{iju}, Y_{ijt}]$ as given in (11.10). Further note that under the normality assumption, one writes

$$E(Y_{ijt} - \pi_{ijt})(Y_{ijv} - \pi_{ijv})(Y_{ijr} - \pi_{ijr}) = 0, \quad (11.33)$$

yielding, by (11.23) – (11.24),

$$\begin{aligned}\delta_{i,jjj,tvr} &= E[Y_{ijt}Y_{ijv}Y_{ijr}] \\ &= \lambda_{ij,tv}\pi_{ijr} + \lambda_{ij,tr}\pi_{ijv} + \lambda_{ij,vr}\pi_{ijt} - 2\pi_{ijt}\pi_{ijv}\pi_{ijr},\end{aligned} \tag{11.34}$$

where, for example, for $t < v$,

$$\lambda_{ij,tv} = \rho^{v-t}\left[\pi_{ijt} - \pi_{ijj,tt}\right] + \pi_{ijj,tv},$$

by (11.24).

Similarly, under the normality assumption, one writes

$$\begin{aligned}& E(Y_{ijt} - \pi_{ijt})(Y_{ijv} - \pi_{ijv})(Y_{ijr} - \pi_{ijr})(Y_{ijd} - \pi_{ijd}) \\ &= \sigma_{i,jj,tv}\sigma_{i,jj,rd} + \sigma_{i,jj,tr}\sigma_{i,jj,vd} + \sigma_{i,jj,td}\sigma_{i,jj,vr},\end{aligned} \tag{11.35}$$

yielding

$$\begin{aligned}\phi_{i,jjjj,tvrd} &= E[Y_{ijt}Y_{ijv}Y_{ijr}Y_{ijd}] \\ &= [\sigma_{i,jj,tv}\sigma_{i,jj,rd} + \sigma_{i,jj,tr}\sigma_{i,jj,vd} + \sigma_{i,jj,td}\sigma_{i,jj,vr} \\ &\quad + \delta_{i,jjj,tvr}\pi_{ijd} + \delta_{i,jjj,tvd}\pi_{ijr} + \delta_{i,jjj,trd}\pi_{ijv} + \delta_{i,jjj,vrd}\pi_{ijt}] \\ &\quad - [\lambda_{ij,tv}\pi_{ijr}\pi_{ijd} + \lambda_{ij,tr}\pi_{ijv}\pi_{ijd} + \lambda_{ij,vr}\pi_{ijt}\pi_{ijd} \\ &\quad + \lambda_{ij,td}\pi_{ijv}\pi_{ijr} + \lambda_{ij,vd}\pi_{ijt}\pi_{ijr} + \lambda_{ij,rd}\pi_{ijt}\pi_{ijv}] \\ &\quad + 3\pi_{ijt}\pi_{ijv}\pi_{ijr}\pi_{ijd}.\end{aligned} \tag{11.36}$$

Note that no extra computation is needed to obtain the elements of the $\mathrm{cov}(U_{ij(s)})$ in the present binary setup. This is because

$$u_{ij(s)} = [y_{ij1}^2, \ldots, y_{ijt}^2, \ldots, y_{ijT}^2]' \equiv [y_{ij1}, \ldots, y_{ijt}, \ldots, y_{ijT}]'$$

yielding

$$\mathrm{cov}(U_{ij(s)}) = \mathrm{cov}[Y_{ij}] = \Sigma_{ijj},$$

the covariance matrix of the original data as in (11.32).

Next, all aforementioned moments up to order three can be used to construct the desired $\mathrm{cov}(U_{ij(s)}, U_{ij(p)})$ matrix, and similarly all moments up to order four can be used to compute the $\mathrm{var}(U_{ij(p)})$ matrix. For example, two general elements of the $\mathrm{cov}(U_{ij(s)}, U_{ij(p)})$ matrix have the formulas

$$\begin{aligned}\mathrm{cov}[Y_{iju}, Y_{ijt}Y_{ijr}] &= \delta_{i,jjj,utr} - \lambda_{ij,tt}\lambda_{ij,tr} \\ \mathrm{cov}[Y_{iju}, Y_{iju}Y_{ijr}] &= \lambda_{ij,ur}[1 - \pi_{iju}],\end{aligned} \tag{11.37}$$

where $\lambda_{ij,tt} = \pi_{ijt}$ and $\lambda_{ij,tr}$ is given by (11.24). Similarly, two general elements of the var$(U_{ij(p)})$ matrix have the formulas

$$\begin{aligned} \mathrm{var}[Y_{ijt}Y_{ijv}] &= \lambda_{ij,tv}[1-\lambda_{ij,tv}] \\ \mathrm{cov}[Y_{ijt}Y_{ijv}, Y_{ijr}Y_{ijd}] &= \phi_{i,jjjj,tvrd} - \lambda_{ij,tv}\lambda_{ij,rd}. \end{aligned} \tag{11.38}$$

This completes the construction of the $\Omega_{ijj}(N)$ matrix for the Ω_{iN} matrix in (11.27).

Construction of $\Omega_{ijk}(N)$ Matrices for Cases when $j \neq k$

Recall that Ω_{ijk} is the $\{T(T+1)/2\} \times \{T(T+1)/2\}$ covariance matrix of $u_{ij} = (u'_{ij(s)}, u'_{ij(p)})'$ and $u_{ik} = (u'_{ik(s)}, u'_{ik(p)})'$ for $j \neq k,\ j,k = 1,\ldots,n_i$. Note that in order to obtain the formulas for various moments up to order four to construct this $\Omega_{ijk}(N)$ matrix for two selected members $j \neq k$, it is appropriate to construct a stacked random vector

$$y^*_{i,jk} = [y'_{ij}, y'_{ik}]'$$

which the under normality assumption follows the $T(T+1)$-dimensional normal vector with mean

$$\tilde{\pi}_{i,jk} = [\pi'_{ij}, \pi'_{ik}]'$$

with

$$\pi_{ij} = [\pi_{ij1}, \ldots, \pi_{ijt}, \ldots, \pi_{ijT}]'$$

and covariance matrix

$$\tilde{\Sigma}_{ijk} = \begin{bmatrix} \Sigma_{ijj} & \Sigma_{ijk} \\ & \Sigma_{ikk} \end{bmatrix} = (\tilde{\sigma}_{i,jk,ut}),\ T(T+1) \times T(T+1), \tag{11.39}$$

where

$$\begin{aligned} \tilde{\sigma}_{i,jk,ut} &= \mathrm{cov}[Y_{iju}, Y_{ikt}] \\ &= E[Y_{iju}Y_{ikt}] - \pi_{iju}\pi_{ikt} \\ &= \int \pi^*_{iju}(\gamma^*_i)\pi^*_{ikt}(\gamma^*_i) g_N(\gamma^*_i|1) d\gamma^*_i - \pi_{iju}\pi_{ikt} \\ &= \pi_{ijk,ut} - \pi_{iju}\pi_{ikt}, \end{aligned} \tag{11.40}$$

by (11.25).

Now define the third− and fourth-order moments for responses from two members as

$$\begin{aligned} \delta_{i,jjk,tvr} &= E[Y_{ijt}Y_{ijv}Y_{ikr}] \\ &= \lambda_{ij,tv}\pi_{ikr} + \pi_{ijk,tr}\pi_{ijv} + \pi_{ijk,vr}\pi_{ijt} - 2\pi_{ijt}\pi_{ijv}\pi_{ikr}, \end{aligned} \tag{11.41}$$

$$\begin{aligned} \delta_{i,jkk,tvr} &= E[Y_{ijt}Y_{ikv}Y_{ikr}] \\ &= \pi_{ijk,tv}\pi_{ijr} + \pi_{ijk,tr}\pi_{ikv} + \lambda_{ik,vr}\pi_{ijt} - 2\pi_{ijt}\pi_{ikv}\pi_{ikr}, \end{aligned} \tag{11.42}$$

and

$$\begin{aligned}\phi_{i,jjkk,tvrd} &= E[Y_{ijt}Y_{ijv}Y_{ikr}Y_{ikd}]\\ &= [\tilde{\sigma}_{i,jj,tv}\tilde{\sigma}_{i,kk,rd} + \tilde{\sigma}_{i,jk,tr}\tilde{\sigma}_{i,jk,vd} + \tilde{\sigma}_{i,jk,td}\tilde{\sigma}_{i,jk,vr}\\ &\quad +\delta_{i,jjk,tvr}\pi_{ijd} + \delta_{i,jjk,tvd}\pi_{ijr} + \delta_{i,jkk,trd}\pi_{ijv} + \delta_{i,jkk,vrd}\pi_{ijt}]\\ &\quad -[\lambda_{ij,tv}\pi_{ijr}\pi_{ijd} + \pi_{ijk,tr}\pi_{ijv}\pi_{ikd} + \pi_{ijk,vr}\pi_{ijt}\pi_{ikd}\\ &\quad +\pi_{ijk,td}\pi_{ijv}\pi_{ikr} + \pi_{ijk,vd}\pi_{ijt}\pi_{ikr} + \lambda_{ik,rd}\pi_{ijt}\pi_{ijv}]\\ &\quad +3\pi_{ijt}\pi_{ijv}\pi_{ikr}\pi_{ikd}. \end{aligned} \tag{11.43}$$

Now by using (11.41) – (11.43), one can compute any element of the $\Omega_{ijk}(N)$ matrix. For example,

$$\mathrm{cov}(Y_{ijt}, Y_{ikv}) = \tilde{\sigma}_{i,jk,tv} \tag{11.44}$$

$$\mathrm{cov}(Y_{ijt}, Y_{ikr}Y_{ikv}) = \delta_{i,jkk,trv} - \pi_{ijt}\lambda_{ik,rv} \tag{11.45}$$

$$\mathrm{cov}(Y_{iju}Y_{ijt}, Y_{ikv}) = \delta_{i,jjk,utv} - \lambda_{ij,ut}\pi_{ikv} \tag{11.46}$$

$$\mathrm{cov}(Y_{ijt}Y_{ijv}, Y_{ikr}Y_{ikd}) = \phi_{i,jjkk,tvrd} - \lambda_{ij,tv}\lambda_{ik,rd}. \tag{11.47}$$

This completes the construction of the $\Omega_{ijk}(N)$ matrices.

11.1.4.3 Moment Estimation of Longitudinal Correlation Index Parameter ρ

Note that the iterative solution of the estimating equation (11.16) for β and (11.26) or (11.27) for σ^2_γ requires a consistent estimator for the longitudinal correlation index parameter ρ. This consistent estimation for ρ may be achieved by using the method of moments. Recall from (11.9) that under the AR(1) process, the lag 1 covariance between y_{ijt} and $y_{ij,t+1}$ is given by

$$\mathrm{cov}(Y_{ijt}, Y_{ij,t+1}) = \rho\,[\pi_{ijt} - \pi_{ijj,tt}] + \left[\pi_{ijj,t,t+1} - \pi_{ijt}\pi_{ij,t+1}\right]. \tag{11.48}$$

For known β and σ^2_γ, one may then obtain the moment estimator of ρ, which is consistent, by equating the sample lag 1 autocovariance with its population counterpart. To be specific, the moment estimator of ρ under the AR(1) process has the formula given by

$$\hat{\rho}_M = \frac{a_1 - b_1}{c_1}, \tag{11.49}$$

where

$$a_1 = \frac{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-1} \tilde{y}_{ijt}\tilde{y}_{ij(t+1)} / \left\{(T-1)\sum_{i=1}^K n_i\right\}}{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T} \tilde{y}^2_{ijt} / \left\{T\sum_{i=1}^K n_i\right\}},$$

$$b_1 = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{t=1}^{T-1}\left[\pi_{ijj,t,t+1} - \pi_{ijt}\pi_{ij,t+1}\right]}{(T-1)\sum_{i=1}^{K} n_i},$$

and

$$c_1 = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{t=1}^{T-1}\left[\pi_{ijt} - \pi_{ijj,tt}\right]}{\left\{(T-1)\sum_{i=1}^{K} n_i\right\}},$$

where

$$\tilde{y}_{ijt} = \frac{(y_{ijt} - \pi_{ijt})}{\sigma_{i,jj,tt}^{\frac{1}{2}}},$$

with $\sigma_{i,jj,tt}(\beta,\sigma_\gamma^2) = \pi_{ijt}(\beta,\sigma_\gamma^2)[1-\pi_{ijt}(\beta,\sigma_\gamma^2)]$ by (11.8).

This correlation estimate from (11.49) is used in (11.16) and (11.26) [or (11.27)] to obtain further improved estimates of β and σ_γ^2, respectively, which are in turn used in (11.49) to obtain further improved estimate of ρ. This cycle of iteration continues until convergence.

11.2 Application to Waterloo Smoking Prevention Data

In the GQL estimation approach discussed in the last section, the regression effects β for the familial longitudinal model were estimated by (11.16), the longitudinal correlation parameter was estimated by (11.49), whereas it was suggested to estimate the familial correlation index parameter σ_γ^2 either by using the $\Omega_i(I)$ based GQL(I) estimating equation (11.26) or by using the Ω_{iN} based GQL(N) estimating equation (11.27). Sutradhar and Farrell (2004), for example, have applied this GQL(I) approach to cluster-correlated binary longitudinal data from the Waterloo Smoking Prevention Project, Study 3. They have, however, used a stationary correlation structure based GQL(I) approach which does not require the specification of a longitudinal correlation structure whether AR(1), MA(1), or EQC, but this stationary correlations based approach is suitable when covariates are time-independent. Sashegyi, Brown, and Farrell (2000) have applied a generalized penalized quasi-likelihood (GPQL) approach to analyze the same smoking prevention study data. This GPQL approach is a generalization of the PQL approach of Breslow and Clayton (1993) from the generalized linear mixed models (GLMMs) setup to the generalized linear longitudinal mixed models (GLLMMs) setup. Note, however, that this GPQL approach appears to have several pitfalls. First, as mentioned above, the best linear unbiased prediction (BLUP) analogue PQL approach does not produce a consistent estimate for the variance component of the random effects for small cluster sizes (see Section 4.2.2 in the context of count data). Second, in this approach, one estimates the conditional correlations by using the unconditional sample correlations, yielding inconsistent estimates.

In the smoking prevention study [See Best et al. (1995); Brown and Cameron (1997); and Sashegyi, Brown, and Farrell (2000)], initially 100 Southern Ontario elementary schools were assigned to either a control condition receiving no intervention or one of the four treatment conditions distinguished by the type of provider delivering the intervention program (nurse or teacher), and the type of training the provider had received (workshop participation or self-preparation through printed material). A baseline measure of smoking status was taken prior to any intervention at the beginning of Grade 6. Subsequent assessments were then made at the end of Grades 7 and 8, after which the students moved on to secondary schools.

As part of the high school component of the study, the students of the elementary cohort were followed to the end of Grade 12, and their smoking status was measured annually for four years starting in Grade 9. In this part of the study, 30 high schools, each of which enrolled 30 or more students from the original cohort, were randomly assigned to either an intervention or control condition.

We focus here on a subset of the data from the secondary school component of the study only. Specifically, we attempted to construct a dataset by selecting a simple random sample of four students at each high school from among those individuals at the institution who provided complete information from Grades 9 through 12. This was not possible for one of the schools due to an insufficient number of students providing complete information. However, appropriate samples of size four were selected from the other 29 schools in order to create the dataset.

In notation, the response y_{ijt} = 1 if the jth student $(j = 1,\ldots,4)$, attending the ith school $(i = 1,\ldots,29)$ was a smoker at time period $t (t = 1,\ldots,4)$, and 0 otherwise. We assume that the random school level effects, $\gamma_i \sim$ i.i.d. $N(0,\sigma_\gamma^2)$. The covariates included in the model were as follows.

- t, grade effect, $t = 1,\ldots,4$, where $t = 1$ represents a Grade 9 observation, and so on.
- HS, high school study condition, $HS = 1$ for intervention schools, and 0 otherwise.
- ES, elementary school study condition, $ES = 1$ for any one of the four types of intervention schools, and 0 otherwise.
- $GENDER$, gender effect, taking 1 for females, and 0 for males.
- $IRISK$, individual level risk score, $IRISK = 1$ for students deemed to be at high risk for smoking based on the habits of parents, siblings, and friends, and 0 otherwise.

In addition, the interaction between elementary and high school conditions HS x ES, and the interaction between high school condition and gender HS x $GENDER$, were also considered. Note that because the covariates are time independent, Sutradhar and Farrell (2004) have considered a stationary correlation structure among the responses over time. Thus, $\mathrm{Corr}(y_{ijt}, y_{iju}|\gamma_i) = \rho_l^* = \rho_{|t-u|}^*$, where $l = |t-u|$, and $l = 1,\ldots,3$. Note that this stationary correlation structure produces the unconditional covariances as

$$\mathrm{cov}(Y_{iju}, Y_{ikt}) = \begin{cases} \rho^*_{|t-u|} g_{ijj,ut} + [\pi_{ijj,ut} - \pi_{iju}\pi_{ijt}] & \text{for } k = j; u < t \\ [\pi_{ijk,ut} - \pi_{iju}\pi_{ikt}] & \text{for } k \neq j; u \leq t, \end{cases} \tag{11.50}$$

where

$$g_{ijj,ut} = \int_{-\infty}^{\infty} [\pi^*_{iju}\{1 - \pi^*_{iju}\}\pi^*_{ijt}\{1 - \pi^*_{ijt}\}]^{1/2} g_N(\gamma^*_i|1) d\gamma^*_i.$$

Now by using $\sigma_{i,jj,ut}$ as the function of $\rho^*_{|t-u|}$ by (11.50), and

$$\lambda_{ij,ut} = \rho^*_{|t-u|} g_{ijj,ut} + \pi_{ijj,ut},$$

the estimating equation (11.16) for β and the GQL(I) estimating equation (11.26) for σ^2_γ are solved for given values of $\rho^*_\ell(\ell = 1, \ldots, 3)$. For the estimation of ρ^*_ℓ, similar to (11.49), we use the method of moments and obtain

$$\hat{\rho}_{\ell,M} = \frac{a^*_1 - b^*_1}{c^*_1}, \tag{11.51}$$

where

$$a^*_1 = \frac{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-\ell} \tilde{y}_{ijt}\tilde{y}_{ij(t+\ell)} / \left\{(T-\ell)\sum_{i=1}^K n_i\right\}}{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T} \tilde{y}^2_{ijt} / \left\{T \sum_{i=1}^K n_i\right\}},$$

$$b^*_1 = \frac{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-\ell} \left[\pi_{ijj,t,t+\ell} - \pi_{ijt}\pi_{ij,t+\ell}\right]}{(T-\ell)\sum_{i=1}^K n_i},$$

and

$$c^*_1 = \frac{\sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^{T-\ell} g_{i,jj,t,t+\ell}}{\left\{(T-\ell)\sum_{i=1}^K n_i\right\}}.$$

The results from fitting the model using both the GQL(I) and GPQL approaches are presented in Table 11.1. There is little difference in the estimates of the fixed effects parameters in the model. Results obtained under both approaches seem to suggest that there is a significant grade effect, and that the individual level risk score also influences smoking status. In addition, they indicate that there appears to be a significant interaction between the high school study condition and gender, regardless of the estimation approach used. There are, however, noticeable differences in the GQL(I) and GPQL estimates of the correlation coefficients and the random effects variance. Also of note is the sizeable difference in the estimated standard errors of the GQL(I) and GPQL estimators of the random effects variance, with the former being much smaller. These results are expected and in agreement with the aforementioned pitfalls in the GPQL approach. For a simulation study on the finite sample relative performances of the GQL(I) and GPQL approaches, revealing a similar pattern in estimates, we refer to Sutradhar and Farrell (2004).

Table 11.1 GQL(I) and GPQL estimates of model parameters and associated standard errors obtained using the subset of the Waterloo Smoking Prevention Project-3 high school data.

TERM	GQL(I) Estimate	GQL(I) Std Err	GPQL Estimate	GPQL Std Err
Intercept	−1.328	0.409	−1.339	0.418
t	0.381	0.089	0.372	0.093
HS	−0.416	0.464	−0.401	0.455
ES	−0.398	0.373	−0.400	0.365
HS x *ES*	0.263	0.490	−0.256	0.482
Gender	−0.348	0.292	−0.360	0.298
HS x *Gender*	1.008	0.415	0.994	0.413
Irisk	0.718	0.224	0.712	0.229
ρ_1^*	0.589	–	0.466	–
ρ_2^*	0.381	–	0.270	–
ρ_3^*	0.233	–	0.145	–
σ_γ^2	1.199	0.032	2.384	0.281

11.3 Family Based BDML Models for Binary Data

In Sections 11.1 and 11.2, we have discussed a class of family based linear dynamic conditional probability models to analyze familial longitudinal binary data. However, as argued in Chapter 9 (see Section 9.2), there are situations when, as opposed to the linear dynamic models, the nonlinear dynamic logit model may explain the binary data well. This happens especially when it is expected that the marginal mean and variance at a given time depend on the past means and variances, respectively, in a recursive way. Also, a technical advantage of this nonlinear logit model over the LDCP models is that the correlations in the nonlinear setup satisfy the full range from -1 to $+1$. In this section, for the purpose of fitting familial longitudinal binary data, we consider a generalization of the BDML (binary dynamic mixed logit) models used in Section 9.2 in the longitudinal setup, to the familial longitudinal setup. This family based new longitudinal mixed model is referred to as the family based BDML (FBDML) model. Because, as opposed to the LDCP models, it is feasible to use the likelihood approach in this logit model setup, we discuss the likelihood estimation approach in addition to the GQL approach for fitting such FBDML models. The GQL and ML (maximum likelihood) approaches are developed in Sections 11.3.2 and 11.3.3, respectively, whereas the basic properties of the FBDML model are given in Section 11.3.1.

11.3.1 FBDML Model and Basic Properties

As opposed to the conditional linear dynamic binary probability model given in (11.1), we consider a generalization of the mixed logit model (9.27) to the familial longitudinal setup. This FBBDML model is given by

$$Pr(Y_{ij1} = 1|x_{ij1};\gamma_i) = \frac{\exp(x'_{ij1}\beta + \gamma_i)}{[1+\exp(x'_{ij1}\beta + \gamma_i)]} = p^*_{ij10}(\gamma_i), \quad (11.52)$$

$$\begin{aligned} Pr(Y_{ijt} = 1|x_{ijt}, y_{ij,t-1};\gamma_i) &= \frac{\exp\{x'_{ijt}\beta + \theta y_{ij,t-1} + \gamma_i\}}{[1+\exp\{x'_{ijt}\beta + \theta y_{ij,t-1} + \gamma_i\}]} \\ &= p^*_{ijty_{ij,t-1}}(\gamma_i), \end{aligned} \quad (11.53)$$

for $t = 2,\ldots,T$. In $(11.52)-(11.53)$, β is the p-dimensional vector of regression effects, θ is the lag 1 dynamic dependence parameter, and γ_i is the unobservable random effect for the ith family. As far as the distribution of γ_i is concerned, similar to the LDCCP models, we assume that $\gamma_i \overset{iid}{\sim} N(0,\sigma^2_\gamma)$ [Breslow and Clayton (1993); Sutradhar (2004)] so that for $\gamma^*_i = \gamma_i/\sigma_\gamma$, $\gamma^*_i \overset{iid}{\sim} N(0,1)$.

11.3.1.1 Conditional Mean, Variance, and Correlation Structures

Note that the familial longitudinal model defined by $(11.52)-(11.53)$ may be treated as a generalization of the binary longitudinal mixed model for an individual considered by Sutradhar, Rao, and Pandit (2008). Further note that the binary dynamic model defined by (11.52) and (11.53) appears to be quite suitable to interpret the data for many health problems. For example, this model produces the mean (also the variance) at a given time point for an individual member of a given family as a function of the covariate history of the individual up to the present time. This history based mean function appears to be useful to interpret the current asthma status (yes or no) of an individual as a function of the related covariates such as smoking habits and cleanliness over a suitable past period. In notation, conditional on the random effect γ_i, the marginal mean at a given point of time t, that is, $\pi^*_{ijt}(\gamma_i) = E[Y_{ijt}|\gamma_i] = Pr(Y_{ijt} = 1|\gamma_i)$, for $t = 2,\ldots,T$, has the dynamic relationship with past means as given by

$$\pi^*_{ijt}(\gamma^*_i) = E[Y_{ijt}|\gamma^*_i] = p^*_{ijt0}(\gamma^*_i) + \pi^*_{ij,t-1}(\gamma^*_i)(p^*_{ijt1}(\gamma_i) - p^*_{ijt0}(\gamma^*_i)), \quad (11.54)$$

with

$$\pi^*_{ij1}(\gamma^*_i) = p^*_{ij10}(\gamma^*_i) = \frac{\exp(x'_{ij1}\beta + \sigma_\gamma\gamma^*_i)}{[1+\exp(x'_{ij1}\beta + \sigma_\gamma\gamma^*_i)]}$$

by (11.52), and for other t, $p^*_{ijt1}(\gamma^*_i)$ and $p^*_{ijt0}(\gamma^*_i)$ are given by (11.53) with $\gamma_i = \sigma_\gamma\gamma^*_i$. By a similar operation as in (11.54), the formula for the covariance between y_{iju} and y_{ijt}, conditional on γ^*_i, may be obtained as

$$\sigma^*_{i,jj,ut}(\gamma^*_i) = \text{cov}(Y_{iju}, Y_{ijt}|\gamma^*_i)$$
$$= \pi^*_{iju}(\gamma^*_i)(1 - \pi^*_{iju}(\gamma^*_i))\Pi^t_{m=u+1}(p^*_{ijm1}(\gamma^*_i) - p^*_{ijm0}(\gamma^*_i)). \quad (11.55)$$

Furthermore, we assume that conditional on γ_i, the responses of any two members are independent irrespective of their recording times. That is,

$$\text{cov}[(Y_{iju}, Y_{ikt})|\gamma_i] = 0, \text{ for } j \neq k. \quad (11.56)$$

It then follows that conditional on γ^*_i, the correlation between y_{iju} and y_{ikt} has the form given by

$$\rho^*_{i,jk,ut}(\gamma^*_i) = \text{corr}\{(Y_{iju}, Y_{ikt})|\gamma^*_i\}$$
$$= \begin{cases} \sqrt{\frac{\pi^*_{iju}(\gamma^*_i)(1-\pi^*_{iju}(\gamma^*_i))}{\pi^*_{ijt}(\gamma^*_i)(1-\pi^*_{ijt}(\gamma^*_i))}}\,\Pi^t_{m=u+1}(p^*_{ijm1}(\gamma^*_i) - p^*_{ijm0}(\gamma^*_i)), & \text{for } j = k \\ 0, & \text{for } j \neq k, \end{cases}$$
$$(11.57)$$

which ranges between -1 and $+1$. Thus, the nonlinear dynamic model $(11.52)-(11.53)$ produces correlations with full ranges, whereas the conditional linear dynamic probability model (11.1) produces correlations with narrower ranges.

It may also be convenient to have the formulas for the second-order raw moments. Conditional on γ^*_i, for $u < t$, these raw second-order expectations, by following (11.55) and (11.56), may be written as

$$E(Y_{iju}Y_{ikt}|\gamma^*_i) = \begin{cases} \sigma^*_{i,jj,ut}(\gamma^*_i) + \pi^*_{iju}(\gamma^*_i)\pi^*_{ijt}(\gamma^*_i) = \lambda^*_{ijjut}, & \text{for } j = k \\ \pi^*_{iju}(\gamma^*_i)\pi^*_{ikt}(\gamma^*_i) = \lambda^*_{ijkut} & \text{for } j \neq k, \end{cases} \quad (11.58)$$

where $\pi^*_{iju}(\gamma^*_i)$ is given by (11.54), and it is quite different from the conditional binary probability in (11.2) under the linear dynamic model.

11.3.1.2 Unconditional Mean, Variance, and Correlation Structures

As far as the unconditional means, variances, and covariances are concerned, they may be derived from $(11.54)-(11.56)$ as follows. First, to compute the unconditional mean from the conditional mean in (11.54), we write

$$E[Y_{ijt}] = E_{\tau_i}E[Y_{ijt}|\gamma^*_i]$$
$$= \int_{-\infty}^{\infty} \pi^*_{ijt}(\gamma^*_i)g_N(\gamma^*_i|1)d\gamma^*_i$$
$$= \pi_{ijt}, \quad (11.59)$$

where the formula for $\pi^*_{ijt}(\gamma^*_i)$ is given in (11.54).

In the fashion similar to that of (11.59), one may next compute the unconditional second-order moments $E[Y_{ijt}^2]$ and $E[Y_{iju}Y_{ikt}]$ by using (11.58), as

$$E(Y_{iju}Y_{ikt}) = \begin{cases} \int_{-\infty}^{\infty} \lambda_{ijjut}^*(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* = \lambda_{ijjut} \text{ for } j=k, \\ \int_{-\infty}^{\infty} \lambda_{ijkut}^*(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^* = \lambda_{ijkut} \text{ for} j \neq k, \end{cases} \tag{11.60}$$

yielding the variances and the covariances as

$$\begin{aligned} \text{var}[Y_{iju}] &= \sigma_{i,jj,uu} = \lambda_{ijjuu} - \pi_{iju}^2 \\ \text{cov}[Y_{iju}, Y_{ikt}] &= \sigma_{i,jk,ut} = \lambda_{ijkut} - \pi_{iju}\pi_{ikt}, \end{aligned} \tag{11.61}$$

respectively. The integrals in (11.59) and (11.60) may be computed by using the binomial approximation as used in (11.7).

Note that the aforementioned first− and second-order unconditional moments along with other higher-order moments up to order four are used in the next section to develop the so-called GQL estimating equations for the regression parameters vector β, dynamic dependence parameter θ, and the variance of the random effects σ_γ^2. The likelihood equations for these parameters are given in Section 11.3.3.

11.3.2 Quasi-Likelihood Estimation in the Familial Longitudinal Setup

11.3.2.1 Joint GQL Estimation of Parameters

Recall that in Sections 11.1 and 11.2, the regression effects β and the random effects variance σ_γ^2 were estimated by solving GQL estimating equations, whereas the longitudinal correlation index parameter ρ was estimated by using the method of moments. Note that in the present FBDML model, the dynamic dependence parameter θ is equivalent to ρ, but, it is more similar to the β parameter when past responses are thought to be certain regression covariates. Consequently, we chose to estimate all three parameters, namely, β, θ, and σ_γ, simultaneously by solving a joint GQL estimating equation.

For

$$\begin{aligned} y_i &= [y_{i11}, \ldots, y_{ijt}, \ldots, y_{in_iT}]' : n_iT \times 1, \\ s_i^* &= [y_{i11}^2, \ldots, y_{ijt}^2, \ldots, y_{in_iT}^2]' : n_iT \times 1, \\ s_{i1} &= [y_{i11}y_{i12}, \ldots, y_{iju}y_{ijt}, \ldots, y_{in_i(T-1)}y_{in_iT}]' : n_iT(T-1)/2 \times 1, \text{ and} \\ s_{i2} &= [y_{i11}y_{i21}, \ldots, y_{iju}y_{ikt}, \ldots, y_{i(n_i-1)T}y_{in_iT}]' : \frac{n_i(n_i-1)}{2}T^2 \times 1, \end{aligned} \tag{11.62}$$

let

$$f_i = [y_i', s_{i1}', s_{i2}']' \tag{11.63}$$

be the vector of distinct first− and second-order responses. Furthermore, let $\alpha = (\beta', \theta, \sigma_\gamma)'$ be the vector of all parameters involved in the model $(11.52)-(11.53)$. Following Sutradhar (2004) [see also Sutradhar rao, and Pandit (2008)], the GQL estimating equations for the components of the α vector may be written as

$$\sum_{i=1}^{K} \frac{\partial \xi_i'}{\partial \alpha} \Omega_i^{-1} (f_i - \xi_i) = 0, \tag{11.64}$$

where

$$\begin{aligned} \xi_i = E(F_i) &= [E(Y_i'), E(S_{i1}', E(S_{i2}')]' \\ &= [\mu_i', \lambda_{i1}', \lambda_{i2}']' \ \ \text{(say)} \end{aligned} \tag{11.65}$$

$$\begin{aligned} \Omega_i = \text{cov}(F_i) &= \begin{bmatrix} \text{cov}(Y_i) & \text{cov}(Y_i, S_{i1}') & \text{cov}(Y_i, S_{i2}') \\ \cdot & \text{cov}(S_{i1}) & \text{cov}(S_{i1}, S_{i2}') \\ \cdot & \cdot & \text{cov}(S_{i2}) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_i & \Delta_{i11} & \Delta_{i12} \\ \cdot & \Omega_{i11} & \Omega_{i12} \\ \cdot & \cdot & \Omega_{i22} \end{bmatrix} \ \ \text{(say).} \end{aligned} \tag{11.66}$$

and $\partial \xi_i'/\partial \alpha$ denotes the matrix of the first derivative of ξ_i with respect to the components of α.

Note that the formula for π_{ijt} is given by (11.59). One may then construct the mean vector π_i as

$$\pi_i = [\pi_{i1}', \ldots, \pi_{ij}', \ldots, \pi_{in_i}']',$$

where $\pi_{ij} = [\pi_{ij1}, \ldots, \pi_{ijt}, \ldots, \pi_{ijT}]'$. Note that even though the mean vector π_i here looks similar to that of (11.16), the formula for its components are, however, quite different. Similarly, the $n_iT \times n_iT$ covariance matrix Σ_i for (11.61) can be computed by using the formulas for the var$[Y_{ijt}]$ and cov$[Y_{iju}, Y_{ikt}]$ from (11.61). As far as the computation of the other mean vectors, namely, λ_{i1} and λ_{i2} in (11.65), and the computation of the other component matrices for Ω_i in (11.66) are concerned, they may be done similarly. However, because the construction of some of the matrices in (11.66) require fourth-order moment computations, for practical convenience, we provide below a simpler uniform and numerically friendly computational technique for all components of the mean vector ξ_i and covariance matrix Ω_i.

(a) Formulas for the Elements of the Mean Vector ξ_i for GQL Estimating Equation (11.64)

Let

$$\tilde{p}^*_{ij10}(\tau_i) = Pr(y_{ij1}|\tau_i) = p^{*y_{ij1}}_{ij10}(\tau_i)(1 - p^*_{ij10}(\tau_i))^{1-y_{ij1}}$$

$$\tilde{p}^*_{ijty_{i,t-1}}(\tau_i) = Pr(y_{ijt}|y_{ij,t-1}, \tau_i) = p^{*y_{ijt}}_{ijty_{ij,t-1}}(\tau_i) \times (1 - p^*_{ijty_{ij,t-1}}(\tau_i))^{1-y_{ijt}}, \quad (11.67)$$

where $p^*_{ijty_{ij,t-1}}$ are defined in $(11.52) - (11.53)$. We now first compute the conditional first-order moment $\pi^*_{ijt}(\gamma^*_i)$ as

$$\pi^*_{ijt}(\gamma^*_i) = E[Y_{ijt}|\gamma^*_i] = \Pi^{n_i}_{j=1}[\tilde{p}^*_{ij10}(\gamma^*_i)] \sum_{(y_{ijt}) \notin s} \Pi^{n_i}_{j=1} \Pi^T_{t=2} \left[\tilde{p}^*_{ijty_{ij,t-1}}(\gamma^*_i)\right]_{(y_{ijt}=1)}, \quad (11.68)$$

where $\sum_{(y_{ijt}) \notin s}$ indicates the summation over all responses in the sample space $'s'$ which in this case does not contain y_{ijt} only. More specifically, here $'s'$ contains all binary responses recorded at all time points for all members except the response from the jth member at the tth time under the ith family. Similarly, for any $j,k = 1,\ldots,n_i$, and $u,t = 1,\ldots,T$, we compute the conditional second-order moments as

$$\lambda^*_{ijkut}(\gamma^*_i) = E[Y_{iju}Y_{ikt}|\gamma^*_i] = \Pi^{n_i}_{j=1}[\tilde{p}^*_{ij10}(\gamma^*_i)] \sum_{(y_{iju}, y_{ikt}) \notin s} \Pi^{n_i}_{j=1} \Pi^T_{t=2} \left[\tilde{p}^*_{ijty_{ij,t-1}}(\gamma^*_i)\right]_{(y_{iju}=1, y_{ikt}=1)}. \quad (11.69)$$

By using (11.68) and (11.69), we now compute the unconditional first− and second-order moments by

$$\pi_{ijt} = E(Y_{ijt}) = \int_{-\infty}^{\infty} \pi^*_{ijt}(\gamma^*_i) g_N(\gamma^*_i|1) d\gamma^*_i \quad (11.70)$$

$$\lambda_{ijkut} = E(Y_{iju}Y_{ikt}) = \int_{-\infty}^{\infty} \lambda^*_{ijkut}(\gamma^*_i) g_N(\gamma^*_i|1) d\gamma^*_i, \quad (11.71)$$

which may be evaluated by using the binomial approximation similar to (11.7). It is then clear that (11.70) leads to the vector π_i and similarly (11.71) leads to the vectors λ_{i1} and λ_{i2}, completing the computation of the mean vector ξ_i for the GQL estimating equation (11.64).

(b) Formulas for the Elements in the Submatrices of the Weight Matrix Ω_i for GQL Estimating Equation (11.64)

Recall that the weight matrix Ω_i in (11.64) contains the submatrices as shown by (11.66). The computation of these submatrices requires the knowledge of the second−, third− and fourth-order moments of the data. More specifically, Σ_i is the second-order moment matrix, which may be computed by using the formulas for the first and second moments given by (11.68) and (11.69). For example, a general formula for the diagonal elements of this $n_iT \times n_iT$ matrix is given by

$$\mathrm{var}(Y_{iju}) = \pi_{iju}(1-\pi_{iju}), \tag{11.72}$$

whereas a general formula for the off-diagonal elements is given by

$$\mathrm{cov}(Y_{iju}, Y_{ikt}) = \lambda_{ijkut} - \pi_{iju}\pi_{ikt}. \tag{11.73}$$

b(i) Computation of the Third-Order Moment Matrices Δ_{i11} and Δ_{i12}

These matrices can be computed by using two different general elements, namely, $\mathrm{cov}(Y_{iju}, Y_{iju}Y_{ikt})$ and $\mathrm{cov}(Y_{iju}, Y_{ikt}Y_{i\ell v})$. The first general element computation can be completed by using the first− and second-order moments from (11.70) and (11.71). To be specific, the formula for this general element is given by

$$\begin{aligned}\mathrm{cov}(Y_{iju}, Y_{iju}Y_{ikt}) &= E[Y_{iju}^2 Y_{ikt}] - \pi_{iju}\lambda_{ijkut} \\ &= E[Y_{iju}Y_{ikt}] - \pi_{iju}\lambda_{ijkut} = \lambda_{ijkut}(1-\pi_{iju}).\end{aligned} \tag{11.74}$$

Next to compute the second general element $\mathrm{cov}(Y_{iju}, Y_{ikt}Y_{i\ell v})$, similar to (11.69), we first write the formula for the conditional third-order moment given by

$$\begin{aligned}\psi^*_{ijk\ell utv}(\gamma_i^*) &= E\left(Y_{iju}Y_{ikt}Y_{i\ell v}|\gamma_i^*\right) \\ &= \Pi_{j=1}^{n_i}[\tilde{p}^*_{ij10}(\gamma_i^*)] \\ &\quad \times \sum_{(y_{iju}, y_{ikt}, y_{i\ell v}) \notin s} \Pi_{j=1}^{n_i}\Pi_{t=2}^{T}\left[\tilde{p}^*_{ijty_{ij,t-1}}(\gamma_i^*)\right]_{(y_{iju}=1, y_{ikt}=1, y_{i\ell v}=1)}\end{aligned} \tag{11.75}$$

yielding the unconditional third-order moment as

$$\psi_{ijk\ell utv} = E\left(Y_{iju}Y_{ikt}Y_{i\ell v}\right) = \int_{-\infty}^{\infty} \psi^*_{ijklutv}(\gamma_i^*) g_N(\gamma_i^*|1) d\gamma_i^*. \tag{11.76}$$

We then use this unconditional third moment and compute the desired general element by

$$\mathrm{cov}(Y_{iju}, Y_{ikt}Y_{i\ell v}) = E\left[Y_{iju}Y_{ikt}Y_{i\ell v}\right] - \pi_{iju}\lambda_{ik\ell tv}$$

$$= \psi_{ijk\ell utv} - \pi_{iju}\lambda_{ik\ell tv}. \quad (11.77)$$

b(ii) Computation of the Fourth-Order Moment Matrices $\Omega_{i11},\ \Omega_{i12}$ and Ω_{i22}

The computation of the elements of these matrices requires the computation for the third− and fourth-order moments, where the formulas for the third-order moments are given by (11.76). For the computation of the fourth-order moments, we first write the formula, similar to (11.75), for the conditional fourth-order moment as

$$\begin{aligned}\phi^*_{ijk\ell mutvq}(\gamma_i^*) &= E\left(Y_{iju}Y_{ikt}Y_{i\ell v}Y_{imq}|\gamma_i^*\right)\\ &= \Pi_{j=1}^{n_i}[\tilde{p}^*_{ij10}(\gamma_i^*)]\\ &\times \sum_{(y_{iju},y_{ikt},y_{i\ell v}),y_{imq}\notin s} \Pi_{j=1}^{n_i}\Pi_{t=2}^{T}\left[\tilde{p}^*_{ijty_{ij,t-1}}(\gamma_i^*)\right]_{(y_{iju}=1,y_{ikt}=1,y_{i\ell v}=1,y_{imq}=1)}\end{aligned} \quad (11.78)$$

yielding the unconditional fourth-order moments given by

$$\phi_{ijk\ell mutvq} = E\left(Y_{iju}Y_{ikt}Y_{i\ell v}Y_{imq}|\gamma_i^*\right) = \int_{-\infty}^{\infty}\phi^*_{ijk\ell mutvq}(\gamma_i^*)g_N(\gamma_i^*|1)d\gamma_i^*. \quad (11.79)$$

Finally, these fourth-order moments can be used to compute the desired elements for these matrices as given by

$$\begin{aligned}\text{cov}[Y_{iju}Y_{ikt},Y_{i\ell v}Y_{imq}] &= E\left[Y_{iju}Y_{ikt}Y_{i\ell v}Y_{imq}\right] - E\left[Y_{iju}Y_{ikt}\right]E\left[Y_{i\ell v}Y_{imq}\right]\\ &= \phi_{ijk\ell mutvq} - \lambda_{ijkut}\lambda_{i\ell mvq},\end{aligned} \quad (11.80)$$

where the formula for the second-order moments λ_{ijkut}, for example, is given by (11.71).

(c) Formulas for the Derivatives $\partial\xi_i'/\partial\alpha$ For (11.64)

To compute these derivatives, it is sufficient to compute the derivatives $\partial\pi_{ijt}/\partial\alpha$ and $\partial\lambda_{ijkut}/\partial\alpha$. They are available in Exercises 11.4 and 11.5.

11.3.2.2 Asymptotic Covariance Matrix of the Joint GQL Estimator

Let $\hat{\alpha}_{GQL}$ denote the solution of (11.64). Because the expectation of the GQL estimating function in the left-hand side of (11.64) is zero, this estimator $\hat{\alpha}_{GQL}$ is consistent for α. The GQL estimator $\hat{\alpha}_{GQL}$ is also expected to be highly efficient because of the fact that the GQL estimating equation (11.64) is constructed by using the inverse of the covariance matrix Ω_i as the weight matrix. Furthermore, under some mild regularity conditions it may be shown that $\hat{\alpha}_{GQL}$ asymptotically ($K \to \infty$) follows a Gaussian distribution with mean α and covariance matrix

$$\left[\sum_{i=1}^{K} \frac{\partial \xi_i'}{\partial \alpha} \Omega_i^{-1} \frac{\partial \xi_i}{\partial \alpha'}\right]^{-1}. \quad (11.81)$$

11.3.3 Likelihood Based Estimation

11.3.3.1 Likelihood Function for the FBDML Model

Recall from (11.67) that

$$\tilde{p}^*_{ij10}(\gamma_i^*) = Pr(y_{ij1}|\gamma_i^*) = p^{*y_{ij1}}_{ij10}(\gamma_i^*)(1 - p^*_{ij10}(\gamma_i^*))^{1-y_{ij1}}$$

$$\tilde{p}^*_{ijty_{i,t-1}}(\gamma_i^*) = Pr(y_{ijt}|y_{ij,t-1}, \gamma_i^*) = p^{*y_{ijt}}_{ijty_{ij,t-1}}(\gamma_i^*)(1 - p^*_{ijty_{ij,t-1}}(\gamma_i^*))^{1-y_{ijt}}.$$

The responses of any two members of the same family at any two time points, say y_{iju} and y_{ikt}, conditional on the family effect γ_i^*, are assumed to be independent, thus it follows that the likelihood function for $\beta,\ \theta,$ and σ_γ is given by

$$L(\beta, \theta, \sigma_\gamma) = \Pi_{i=1}^{K} \int_{-\infty}^{\infty} \Pi_{j=1}^{n_i} \left[\tilde{p}^*_{ij10}(\gamma_i^*) \Pi_{t=2}^{T} \tilde{p}^*_{ijty_{ij,t-1}}(\gamma_i^*)\right] \phi(\gamma_i^*) d\gamma_i^*, \quad (11.82)$$

where $\phi(\gamma_i^*)$, for example, is the standard normal density of γ_i^*. This leads to the log-likelihood function as

$$\log L(\beta, \theta, \sigma_\gamma) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{t=1}^{T} \left[y_{ijt} x_{ijt}' \beta + \theta y_{ijt} y_{ij,t-1}\right] + \sum_{i=1}^{K} \log J_i, \quad (11.83)$$

where, for technical convenience, we assume $y_{ij0} = 0$ for all $j = 1, \ldots, n_i$, and $i = 1, \ldots, K$, and J_i has the form given by

$$J_i = \int_{-\infty}^{\infty} \exp(d_i s_i) \Delta_i \phi(\tau_i) d\tau_i,$$

with

$$s_i = \sum_{j=1}^{n_i} \sum_{t=1}^{T} y_{ijt}, \quad d_i = \sigma_\gamma \gamma_i^*,$$

and

$$\Delta_i = \left[\Pi_{j=1}^{n_i} \Pi_{t=1}^{T} \{1 + \exp(x_{ijt}' \beta + \theta y_{ij,t-1} + \sigma_\gamma \gamma_i^*)\}\right]^{-1}.$$

This log-likelihood function is exploited in the next subsection to obtain the likelihood estimates for all three parameters $\beta, \theta,$ and σ_γ.

11.3.3.2 Likelihood Estimating Equations

The likelihood estimating equations for the parameters β, θ, and σ_γ, are obtained by equating their respective score functions to zero. These equations are given by

$$U_1(\beta,\theta,\sigma_\gamma) = \frac{\partial \log L}{\partial \beta} = 0, \quad U_2(\beta,\theta,\sigma_\gamma) = \frac{\partial \log L}{\partial \theta} = 0,$$

$$U_3(\beta,\gamma,\sigma_\tau) = \frac{\partial \log L}{\partial \sigma_\gamma} = 0, \tag{11.84}$$

where the score functions are computed as

$$U_1(\beta,\theta,\sigma_\gamma) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{t=1}^{T} y_{ijt}x_{ijt} - \sum_{i=1}^{K}\frac{A_i}{J_i}, \tag{11.85}$$

$$U_2(\beta,\theta,\sigma_\gamma) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{t=1}^{T} y_{ijt}y_{ij,t-1} - \sum_{i=1}^{K}\frac{B_i}{J_i}, \tag{11.86}$$

and

$$U_3(\beta,\gamma,\sigma_\tau) = \sum_{i=1}^{K}\frac{M_i}{J_i}, \tag{11.87}$$

respectively, where

$$A_i = \int_{-\infty}^{\infty} \exp(d_i s_i)\Delta_i \left[\sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} x_{ijt}\right] \phi(\gamma_i^*)d\gamma_i^*,$$

$$M_i = \int_{-\infty}^{\infty} \exp(d_i s_i)\Delta_i \left[s_i - \sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}}\right] \gamma_i^* \phi(\gamma_i^*)d\gamma_i^*,$$

with $p^*_{ijty_{ij,t-1}}$ as in (11.67). Also, B_i in (11.86) is obtained from A_i by replacing x_{ijt} with $y_{ij,t-1}$. In notation,

$$B_i = A_i|_{x_{ijt} \to y_{ij,t-1}}.$$

The maximum likelihood (ML) estimator of $\alpha = (\beta', \theta, \sigma_\gamma)'$ is then obtained by solving these score equations in (11.84).

Note that the evaluation of J_i, A_i, B_i, and M_i, require an integration, which appears to be quite difficult to solve. As in the last section [see (11.59) for example], we approximate them numerically by using the simulation approach [Jiang (1998); Fahrmeir and Tutz (1994, Chapter 7)]. Note that alternatively, one may use the binomial approximation as in (11.7). For example, we approximate A_i in (11.85) by, say $A_i^{(s)}$ as given by

$$A_i^{(s)} = \frac{1}{M}\sum_{w=1}^{M} \exp(d_{iw}s_i)\Delta_{iw} \left[\sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}}(\gamma^*_{iw})x_{ijt}\right], \tag{11.88}$$

where by (11.83)

$$d_{iw} = \sigma_\gamma \gamma^*_{iw}, \text{ and } \Delta_{iw} = \left[\Pi_{j=1}^{n_i} \Pi_{t=1}^{T} \{1 + \exp(x'_{ijt}\beta + \theta y_{it-1} + \sigma_\gamma \gamma^*_{iw})\}\right]^{-1},$$

and

$$p^*_{ijty_{ij,t-1}}(\gamma^*_{iw}) = \frac{\exp(x'_{ijt}\beta + \theta y_{ij,t-1} + d_{iw})}{1 + \exp(x'_{ijt}\beta + \theta y_{ij,t-1} + d_{iw})},$$

by (11.53). With regard to the numerical approximations, we further remark that one may also achieve them by using other techniques as opposed to the aforementioned simulation approach. For example, one may refer to the adaptive Gaussian quadrature method [Liu and Pierce (1994)] or the so-called binomial approximation [Ten Have and Morabia (1999, eqn. 7)] to the normal integrals. For example, in the later binomial approximation approach, as opposed to the simulation approximation (11.88), one would approximate A_i by $A_i^{(b)}$ (say), similar to (11.7), where for a known reasonably big V such as $V = 5$, and $v_i \sim \text{binomial}(V, \frac{1}{2})$, $A_i^{(b)}$ has the form

$$A_i^{(b)} = \sum_{v_i=0}^{V} \exp(d_i(v_i)s_i)\Delta_i(v_i) \left[\sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}}(v_i)x_{ijt}\right] \binom{V}{v_i} (\frac{1}{2})^{v_i} (\frac{1}{2})^{V-v_i}, \tag{11.89}$$

where, for example,

$$\Delta_i(v_i) = \Delta_i|_{[\gamma_i^* = (v_i - V(1/2))/(V(1/2)(1/2))]},$$

with Δ_i given in (11.83) as a function of γ_i^*.

11.3.3.3 Asymptotic Covariance of the Joint ML Estimator

Let $\alpha = (\beta', \theta, \sigma_\gamma)'$ be the $(p+2)$-dimensional vector of parameters, and $U = [U_1' \; U_2, \; U_3]'$ be the vector of corresponding score functions, where score functions are given by (11.84). Furthermore, let $M = -\partial U/\partial \alpha'$ be the $(p+2) \times (p+2)$ Hessian matrix, which is computed as

$$M = -\begin{bmatrix} \frac{\partial^2 \log L}{\partial\beta\partial\beta'} & \frac{\partial^2 \log L}{\partial\beta\partial\theta} & \frac{\partial^2 \log L}{\partial\beta\partial\sigma_\gamma} \\ \cdot & \frac{\partial^2 \log L}{\partial\theta^2} & \frac{\partial^2 \log L}{\partial\theta\partial\sigma_\gamma} \\ \cdot & \cdot & \frac{\partial^2 \log L}{\partial\sigma_\gamma^2} \end{bmatrix}. \tag{11.90}$$

One may then obtain the maximum likelihood estimate of α by using the iterative

$$\hat{\alpha}_{r+1} = \hat{\alpha}_r + \left[\{E_y M\}^{-1} U\right]_{(r)}, \tag{11.91}$$

where $[\cdot]$ indicates that the quantity in the square bracket is evaluated at $\alpha = \hat{\alpha}_r$ obtained from the rth iteration.

Note that in (11.91), $I(\alpha) = -E_y M$ is known as the Fisher information matrix. However, it is well known that the use of the observed Hessian matrix M in place of $E_y M$ produces almost accurate ML estimate [Efron and Hinkley (1978)]. Thus, instead of (11.91), one may use the Hessian matrix based iterative equation

$$\hat{\alpha}_{r+1} = \hat{\alpha}_r + \left[\{M\}^{-1} U\right]_{(r)}, \tag{11.92}$$

to obtain the ML estimate of α. Let $\hat{\alpha}_{ML}$ be the ML estimator of α which is the solution of (11.92). It also follows that as $K \to \infty$, $\hat{\alpha}_{ML}$ follows the $(p+2)$-dimensional Gaussian distribution with mean α and covariance matrix

$$\text{cov}(\hat{\alpha}_{ML}) = I^{-1}(\alpha) = -[E_y M]^{-1}, \tag{11.93}$$

with its diagonal elements as the asymptotic variances of the ML estimators. Note that an estimate of this asymptotic covariance matrix in (11.93) may be obtained by using the observed Hessian matrix; that is,

$$\hat{\text{cov}}(\hat{\alpha}_{ML}) = \hat{I}^{-1}(\alpha) = -[M]^{-1}. \tag{11.94}$$

However, to examine the efficiency performance of the likelihood approach as compared to the GQL approach discussed in the last section, without doing any expensive simulation study, we need to compute the asymptotic covariance matrix itself given by (11.93). For this purpose we compute the $E_y(M)$ as follows.

Computation of $E_y(M)$

We first compute all six second-order derivatives as shown in Exercise 11.6. These formulas involve the binary responses $y_{ij1}, \ldots, y_{ijT}$ for $j = 1, \ldots, n_i$ and $i = 1, \ldots, K$. Now to obtain the expected values of these derivatives over all possible values of the responses, we first rewrite them by replacing, for example, J_i with J_{iy}, and then taking the sum of the whole derivative function under the ith family over all possible values of the binary responses. For convenience, we demonstrate how to apply this technique to compute the expectation of one of the second-order derivatives, namely for $E\left[\partial^2 \log L/[\partial\beta\partial\beta']\right]$. We re-express the formula for this derivative from Exercise 11.6 as

$$\frac{\partial^2 \log L}{\partial\beta\partial\beta'} = -\sum_{i=1}^{K} \frac{1}{J_{iy}^2}\left[J_{iy} A_{iy\beta} + A_{iy} A'_{iy}\right],$$

and compute its expectation as

$$E\left[\frac{\partial^2 \log L}{\partial\beta\partial\beta'}\right] = -\sum_{i=1}^{K}\left[\sum_{y_{i11}=0}^{1} \cdots \sum_{y_{ijt}=0}^{1} \cdots \sum_{y_{in_iT}=0}^{1} \frac{1}{J_{iy}^2}\left[J_{iy} A_{iy\beta} + A_{iy} A'_{iy}\right]\right]. \tag{11.95}$$

Note that as we are using the simulation (or binomial) approach to approximate A_i by $A_i^{(s)}$, for example, we in fact approximate the expectation in (11.95) as

$$E\left[\frac{\partial^2 \log L}{\partial\beta\partial\beta'}\right] \equiv -\sum_{i=1}^{K}\left[\sum_{yi11=0}^{1}\cdots\sum_{y_{ijt}=0}^{1}\cdots\sum_{y_{in_iT}=0}^{1}\frac{1}{J^{(s)2}_{iy}}\left[J^{(s)}_{iy}A^{(s)}_{iy\beta}+A^{(s)}_{iy}A^{(s)\prime}_{iy}\right]\right]. \quad (11.96)$$

The approximate expectation for five other derivatives may be computed similarly.

Exercises

11.1. (Section 11.1.4.2.) [First-order derivatives of λ_i with respect to σ_γ^2 (equations (11.26) and (11.27))]
Recall the definition of the elements of λ_i from (11.22) to (11.25). Following the notation in (11.25), verify that the derivatives of these elements with respect to σ_γ^2 are obtained by using the following formulas:

$$\frac{\partial\pi_{iju}}{\partial\sigma_\gamma^2} = \frac{1}{2\sigma_\gamma}\sum_{v_i=0}^{V}\left[\frac{v_i-V(1/2)}{V(1/2)(1/2)}\right][\pi^*_{iju}(v_i)\{1-\pi^*_{iju}(v_i)\}]\binom{V}{v_i}(1/2)^{v_i}(1/2)^{V-v_i}$$

$$\frac{\partial\pi_{i,jj,uu}}{\partial\sigma_\gamma^2} = \frac{1}{\sigma_\gamma}\sum_{v_i=0}^{V}\left[\frac{v_i-V(1/2)}{V(1/2)(1/2)}\right]\pi^*_{iju}(v_i)[\pi^*_{iju}(v_i)\{1-\pi^*_{iju}(v_i)\}]$$
$$\binom{V}{v_i}(1/2)^{v_i}(1/2)^{V-v_i}$$

$$\frac{\partial\pi_{i,jj,ut}}{\partial\sigma_\gamma^2} = \frac{1}{2\sigma_\gamma}\sum_{v_i=0}^{V}\left[\frac{v_i-V(1/2)}{V(1/2)(1/2)}\right][\pi^*_{iju}(v_i)\pi^*_{ijt}(v_i)\{2-\pi^*_{iju}(v_i)-\pi^*_{ijt}(v_i)\}]$$
$$\times\binom{V}{v_i}(1/2)^{v_i}(1/2)^{V-v_i},$$

where

$$\pi^*_{ijt}(v_i) = \pi^*_{ijt}(\gamma_i^*)|_{\gamma_i^*=(v_i-V(1/2))/[V(1/2)(1/2)]},$$

γ_i^* being a standard normal variable.

11.2. (Section 11.1.4.2) [Binary marginal and product moments up to order four when $\rho = 0$ for any member(11.29)]
Verify the following results.

$$(a(i))\ \ \text{var}(Y_{ijt}) = \sigma_{ij,tt} = \pi_{ijt}[1-\pi_{ijt}] \text{ by (11.8).} \quad (11.97)$$

$$(a(ii))\ \ \text{cov}(Y_{iju},Y_{ijt}) = \sigma_{ij,ut}|_{\rho=0} \quad (11.98)$$
$$= E_{\gamma_i^*}[\pi^*_{iju}\pi^*_{ijt}] - \pi_{iju}\pi_{ijt}$$
$$= \pi_{ijj,ut} - \pi_{iju}\pi_{ijt} \text{ by (11.10).} \quad (11.99)$$

$$(a(iii))\ \ \text{cov}(Y_{iju},Y_{ijt}Y_{ij\ell}) = E[Y_{iju}Y_{ijt}Y_{ij\ell}] - \pi_{iju}\pi_{ijj,ut}$$
$$= E_{\gamma_i^*}[\pi^*_{iju}\pi^*_{ijt}\pi^*_{ij\ell}] - \pi_{iju}\pi_{ijj,ut}$$

$$= \pi_{ij,ut\ell} - \pi_{iju}\pi_{ijj,ut} \tag{11.100}$$

$$(a(iv))\ \mathrm{var}[Y_{iju}Y_{ijt}] = \pi_{ijj,ut}[1-\pi_{ijj,ut}] \tag{11.101}$$

$$\begin{aligned}(a(v))\ \mathrm{cov}(Y_{iju}Y_{ijt}, Y_{ij\ell}Y_{ijv}) &= E[Y_{iju}Y_{ijt}Y_{ij\ell}Y_{ijv}] - \pi_{ijj,ut}\pi_{ijj,\ell v} \\ &= E_{\gamma_i^*}[\pi^*_{iju}\pi^*_{ijt}\pi^*_{ij\ell}\pi^*_{ijv}] - \pi_{ijj,ut}\pi_{ijj,\ell v} \\ &= \pi_{ijj,ut\ell v} - \pi_{ijj,ut}\pi_{ijj,\ell v}. \end{aligned} \tag{11.102}$$

11.3. (Section 11.1.4.2) [Binary product moments up to order four when $\rho = 0$ for any two members (11.30)]
Verify the following results.

$$\begin{aligned}(b(i))\ \mathrm{cov}(Y_{iju}, Y_{ikt}) &= E[Y_{iju}Y_{ikt}] - \pi_{iju}\pi_{ikt} \\ &= E_{\gamma_i^*}[\pi^*_{iju}\pi^*_{ikt}] - \pi_{iju}\pi_{ikt} \\ &= \pi_{ijk,ut} - \pi_{iju}\pi_{ijt} \end{aligned} \tag{11.103}$$

$$(b(ii))\ \mathrm{cov}(Y_{iju}, Y_{ikt}Y_{ik\ell}) = \pi_{ijk,ut\ell} - \pi_{iju}\pi_{ikk,t\ell} \tag{11.104}$$

$$\begin{aligned}(b(iii))\ \mathrm{cov}(Y_{iju}Y_{ijt}, Y_{ik\ell}Y_{ikv}) &= E[Y_{iju}Y_{ijt}Y_{ik\ell}Y_{ikv}] - \pi_{ijj,ut}\pi_{ikk,\ell v} \\ &= E_{\gamma_i^*}[\pi^*_{iju}\pi^*_{ijt}\pi^*_{ik\ell}\pi^*_{ikv}] - \pi_{ijj,ut}\pi_{ikk,\ell v} \\ &= \pi_{ijk,ut\ell v} - \pi_{ijj,ut}\pi_{ikk,\ell v}. \end{aligned} \tag{11.105}$$

11.4. (Section 11.3.2.1) [Formula for $\partial\pi_{ijt}/\partial\alpha$]
Write $q_i(y_i|\gamma_i^*) = \Pi_{j=1}^{n_i}[\tilde{p}^*_{ij10}(\gamma_i^*)]\Pi_{j=1}^{n_i}\Pi_{t=2}^{T}\left[\tilde{p}^*_{ijty_{ij,t-1}}(\gamma_i^*)\right]$, and show from (11.68) that

$$\frac{\partial\pi^*_{ijt}(\gamma_i^*)}{\partial\beta} = \sum_{y_{ijt}\notin s}\left[q_i(y_i|\gamma_i^*)\sum_{j=1}^{n_i}\sum_{t=1}^{T}(y_{ijt} - p^*_{ijty_{ij,t-1}}(\gamma_i^*))x_{ijt}\right]_{(y_{ijt}=1)}$$

$$\frac{\partial\pi^*_{ijt}(\gamma_i^*)}{\partial\theta} = \sum_{y_{ijt}\notin s}\left[q_i(y_i|\gamma_i^*)\sum_{j=1}^{n_i}\sum_{t=1}^{T}(y_{ijt} - p^*_{ijty_{ij,t-1}}(\gamma_i^*))y_{ij,t-1}\right]_{(y_{ijt}=1)}$$

$$\frac{\partial\pi^*_{ijt}(\gamma_i^*)}{\partial\sigma_\gamma} = \sum_{y_{ijt}\notin s}\left[q_i(y_i|\gamma_i^*)\sum_{j=1}^{n_i}\sum_{t=1}^{T}(y_{ijt} - p^*_{ijty_{ij,t-1}}(\gamma_i^*))\tau_i\right]_{(y_{ijt}=1)}, \tag{11.106}$$

with

$$p^*_{ij1y_{ij0}}(\gamma_i^*) = p^*_{ij10} = \frac{\exp[x'_{ijt}\beta + \sigma_\gamma\gamma_i^*]}{1+\exp[x'_{ijt}\beta + \sigma_\gamma\gamma_i^*]},$$

yielding the desired derivatives as

$$\begin{aligned}\frac{\partial\pi_{ijt}}{\partial\beta} &= \int_{-\infty}^{\infty}\frac{\partial\pi^*_{ijt}(\gamma_i^*)}{\partial\beta}g_N(\gamma_i^*|1)d\gamma_i^* \\ \frac{\partial\pi_{ijt}}{\partial\theta} &= \int_{-\infty}^{\infty}\frac{\partial\pi^*_{ijt}(\gamma_i^*)}{\partial\theta}g_N(\gamma_i^*|1)d\gamma_i^*\end{aligned}$$

$$\frac{\partial \pi_{ijt}}{\partial \sigma_\gamma} = \int_{-\infty}^{\infty} \frac{\partial \pi^*_{ijt}(\gamma^*_i)}{\partial \sigma_\gamma} g_N(\gamma^*_i|1)d\gamma^*_i,$$

respectively.

11.5. (Section 11.3.2.1) [Formula for $\partial \lambda_{ijkut}/\partial \alpha$]

Similar to Exercise 11.4, verify that

$$\frac{\partial \lambda^*_{ijkut}(\gamma^*_i)}{\partial \beta} = \sum_{y_{iju}, y_{ijt} \notin s} \left[q_i(y_i|\gamma^*_i) \sum_{j=1}^{n_i} \sum_{t=1}^{T} (y_{ijt} - p^*_{ijty_{ij,t-1}}(\gamma^*_i)) x_{ijt} \right]_{(y_{iju}=1, y_{ijt}=1)}$$

$$\frac{\partial \lambda^*_{ijkut}(\gamma^*_i)}{\partial \theta} = \sum_{y_{iju}, y_{ijt} \notin s} \left[q_i(y_i|\gamma^*_i) \sum_{j=1}^{n_i} \sum_{t=1}^{T} (y_{ijt} - p^*_{ijty_{ij,t-1}}(\tau_i)) y_{ij,t-1} \right]_{(y_{iju}=1, y_{ijt}=1)}$$

$$\frac{\partial \lambda^*_{ijkut}(\gamma^*_i)}{\partial \sigma_\gamma} = \sum_{y_{iju}, y_{ijt} \notin s} \left[q_i(y_i|\gamma^*_i) \sum_{j=1}^{n_i} \sum_{t=1}^{T} (y_{ijt} - p^*_{ijty_{ij,t-1}}(\gamma^*_i)) \gamma^*_i \right]_{(y_{iju}=1, y_{ijt}=1)} \quad (11{,}107)$$

yielding the desired derivatives as

$$\frac{\partial \lambda_{ijkut}}{\partial \beta} = \int_{-\infty}^{\infty} \frac{\partial \lambda^*_{ijkut}(\gamma^*_i)}{\partial \beta} g_N(\gamma^*_i|1)d\gamma^*_i$$

$$\frac{\partial \lambda_{ijkut}}{\partial \theta} = \int_{-\infty}^{\infty} \frac{\partial \lambda^*_{ijkut}(\gamma^*_i)}{\partial \theta} g_N(\gamma^*_i|1)d\gamma^*_i$$

$$\frac{\partial \lambda_{ijkut}}{\partial \sigma_\gamma} = \int_{-\infty}^{\infty} \frac{\partial \lambda_{ijkut}(\gamma^*_i)}{\partial \sigma_\gamma} g_N(\gamma^*_i|1)d\gamma^*_i,$$

respectively.

11.6. (Section 11.3.3.3) [Formulas for the elements of the second-order derivatives (11.90) of the likelihood function]
Verify that

$$(\mathbf{1}) \quad \frac{\partial^2 \log L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{K} \frac{1}{J_i^2} \left[J_i A_{i\beta} + A_i A'_i \right], \quad (11.108)$$

where J_i and A_i are as in (11.85), and the new quantity $A_{i\beta}$ has the formula

$$A_{i\beta} = \int_{-\infty}^{\infty} \exp(d_i s_i) \Delta_i \left[\{ \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} (1 - p^*_{ijty_{ij,t-1}}) x_{ijt} x'_{ijt} \} - \{ \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} x_{ijt} \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} x'_{it} \} \right] \phi(\gamma^*_i) d\gamma^*_i. \quad (11.109)$$

Similarly

$$(\mathbf{2}) \quad \frac{\partial^2 \log L}{\partial \beta \partial \theta} = -\sum_{i=1}^{K} \frac{1}{J_i^2} \left[J_i A_{i\theta} + A_i B_i \right], \tag{11.110}$$

where the new quantity $A_{i\theta}$ has the formula

$$A_{i\theta} = \int_{-\infty}^{\infty} \exp(d_i s_i) \Delta_i \left[\{ \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} (1 - p^*_{ijty_{ij,t-1}}) x_{ijt} y_{ij,t-1} \} \right.$$

$$\left. - \{ \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} x_{ijt} \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} y_{ij,t-1} \} \right] \phi(\gamma_i^*) d\gamma_i^*. \tag{11.111}$$

Further verify that

$$(\mathbf{3}) \quad \frac{\partial^2 \log L}{\partial \beta \partial \sigma_\gamma} = -\sum_{i=1}^{K} \frac{1}{J_i^2} \left[J_i A_{i\sigma_\gamma} - A_i M_i \right], \tag{11.112}$$

where

$$A_{i\sigma_\gamma} = \int_{-\infty}^{\infty} \exp(d_i s_i) \Delta_i \left[\sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} (1 - p^*_{ijty_{ij,t-1}}) x_{ijt} \right.$$

$$\left. + (\sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} x_{ijt}) (s_i - \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}}) \right]$$

$$\times \gamma_i^* \phi(\gamma_i^*) d\gamma_i^*. \tag{11.113}$$

Next show that the second derivatives with respect to θ and σ_γ, have the formulas given by

$$(\mathbf{4}) \quad \frac{\partial^2 \log L}{\partial \theta^2} = -\sum_{i=1}^{K} \frac{1}{J_i^2} \left[J_i B_{i\theta} + B_i^2 \right], \tag{11.114}$$

$$(\mathbf{5}) \quad \frac{\partial^2 \log L}{\partial \theta \partial \sigma_\gamma} = -\sum_{i=1}^{K} \frac{1}{J_i^2} \left[J_i B_{i\sigma_\gamma} - B_i M_i \right], \tag{11.115}$$

where

$$B_{i\theta} = \int_{-\infty}^{\infty} \exp(d_i s_i) \Delta_i \left[\{ \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} (1 - p^*_{ijty_{ij,t-1}}) y^2_{ij,t-1} \} \right.$$

$$\left. + \{ \sum_{j=1}^{n_i} \sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} y_{ij,t-1} \}^2 \right] \phi(\gamma_i^*) d\gamma_i^*, \tag{11.116}$$

$$B_{i\sigma_\gamma} = \int_{-\infty}^{\infty} \exp(d_i s_i)\Delta_i \left[\{\sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}}(1-p^*_{ijty_{ij,t-1}})y_{ij,t-1}\} \right.$$

$$\left. + \{\sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}} y_{ij,t-1}\}(s_i - \sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}}) \right]$$

$$\times \gamma_i^* \phi(\gamma_i^*) d\gamma_i^*. \tag{11.117}$$

Further show that

$$(\mathbf{6}) \quad \frac{\partial^2 \log L}{\partial \sigma_\gamma^2} = \sum_{i=1}^{K} \frac{1}{J_i^2}\left[J_i M_{i\sigma_\tau} - M_i^2\right], \tag{11.118}$$

where

$$M_{i\sigma_\gamma} = \int_{-\infty}^{\infty} \exp(d_i s_i)\Delta_i \left[(s_i - \sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}})^2 \right.$$

$$\left. - \{\sum_{j=1}^{n_i}\sum_{t=1}^{T} p^*_{ijty_{ij,t-1}}(1-p^*_{ijty_{ij,t-1}})\} \right] \gamma^{*2}_i \phi(\gamma_i^*) d\gamma_i^*. \tag{11.119}$$

References

1. Best, J. A, Brown, K. S., Cameron, R., Manske, S. M., & Santi, S. (1995). Gender and predisposing attributes as predictors of smoking onset: Implications for theory and practice. *J. Health Educ.,* 26, $S52--S60$.
2. Brown, K. S. & Cameron, R. (1997). Long-term evaluation of an elementary and secondary school smoking intervention. Final Report to the National Health Research and Development Program (Canada).
3. Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. . *J. Amer. Statist. Assoc.,* 88, 9 – 25.
4. Breslow, N. E. & Lin, X. (1995). Bias correction in generalized linear models with a single component of dispersion. *Biometrika,* 82, 81 – 92.
5. Efron, B. & Hinkley, D. V. (1978). Assesing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information (with discussion). *Biometrika,* 65, 457 – 487.
6. Fahrmeier, L. & Tutz, G. T.(1994): *Multivariate Statistical Modelling Based on Generalized Linear Models,* New York: Springer-Verlag.
7. Jiang, J. (1998). Consistent estimators in generalized linear mixed models. J. Amer. Statist. Assoc., 93, 720 – 729.
8. Liu, Q. & Pierece, D. (1994). A Note on Gaussian-Hermite Quadrature. *Biometrika,* 81, 624 – 629.
9. Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.
10. Sashegyi, A. I., Brown, K. S. & Farrell, P. J. (2000). Estimation in an empirical Bayes model for longitudinal and cross-sectionally clustered binary data. *Canad. J. Statist.,* 28, 45 – 64.
11. Sutradhar, B. C. (2004). On Exact Quasilikelihood Inference in Generalized Linear Mixed Models. *Sankhya B,* 66, 261 – 289.

12. Sutradhar, B. C. (2010). Inferences in generalized linear longitudinal mixed models. *Canad. J. Statist.,* 38, 174 – 196.
13. Sutradhar, B. C., Rao, R. P., & Pandit, V. N. (2008). Generalized method of moments versus generalized quasi-likelihood inferences in binary panel data models. *Sankhya B,* 70, 34 – 62.
14. Sutradhar, B. C. & Farrell, P. J. (2004). Analyzing multivariate longitudinal binary data: A generalized estimating equations approach. *Canad. J. Statist.,* 32, 39 – 55.
15. Ten Have, T. R. & Morabia, A. (1999). Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometrics,* 55, 85 – 93.

Index

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8,